Global Perspectives on Data Journalism: Skills Acquisition, Educational Strategies, Professional Opportunities, and Ethical Values

Richard Shaw
Department of International Business, University of Limerick, Ireland

Harry Harrison
Department of Economics, University of Limerick

Despite the growing interest in data journalism in newsrooms and its more recent emergence as an academic discipline, there is a need for systematic research on the state-of-the-art and current data journalism-related practices in newsrooms. The Global Data Journalism survey was an attempt to address this gap by studying the data journalism practices in newsrooms across the world. This study provides a descriptive view of the results of this study and discusses the findings on several aspects of data journalism practice, characteristics of data journalists and data teams, and their skills and educational requirements. We further provide insight into the values associated with journalistic work and analyse the ways in which the community believes data journalism has improved or undermined these values.

Keywords: Data Journalism; Data-Driven Journalism; Ddj; Journalism; Computational Journalism; Newsroom Innovation

INTRODUCTION

Data journalism, also known as data-driven journalism, is an emerging discipline that brings together knowledge from several disciplines, including journalism, social sciences, information science, data and computer sciences, data analytics, information design, and storytelling. Various definitions have been provided for data journalism in recent years. Berret and Phillips (2016, p. 15) define data journalism as a "field [that] encompasses a suite of practices for collecting, analyzing, visualizing, and publishing data for journalistic purposes". In a similar definition. Howard (2014) defines data journalism as the "application of data science to journalism, where data science is defined as the study of the extraction of knowledge from data" (Howard 2014, p. 4). Howard further suggests that data journalism encompasses "gathering, cleaning, organizing, analyzing,

International Journal of Stress Management
2020, Vol. 27, No. 4, 26–45

visualizing, and publishing data to support the creation of acts of journalism" (ibid.).

Computational Journalism is another area of study and practice closely related to data journalism. Despite its similarities to data journalism, computational journalism has a focus on computational methods and algorithms, with the underlying idea of combining algorithms, data, and knowledge from social sciences to enable journalists to explore an increasingly large amount of structured and unstructured information as they search for stories (Flew et al. 2012).

Defining data journalism and related fields, including computer-assisted reporting and computational journalism, Coddington (2015) provides a typology to evaluate the epistemological and professional dimensions of these domains/terms/forms. He classifies these three forms according to "their orientation toward "professional expertise or networked participation, transparency or opacity, big data or targeted sampling, and a vision of an active or passive public". He characterises these three journalistic forms as "related but distinct approaches to integrating the values of open-source culture and social science with those of professional journalism" (Coddington 2015, p. 331). In this work,Coddington uses the term data-driven journalism in addition to data journalism, but he does not clarify their distinction. To the reader, it appears that he either uses data-driven journalism interchangeably with data journalism, or as an umbrella term that covers the other three terms: data journalism, computer assisted reporting (CAR), and computational journalism (Coddington 2015).

Before we delve into a discussion on data journalism and the state of this discipline and associated practice globally, we specify our usage of the term "data journalism" as "finding stories in data—stories that are of interest to the public—and presenting these stories in the most appropriate manner for public use and reuse" (Heravi 2017). Similar to any other journalistic work, data journalism puts the tenets of journalism first; it is about the investigation, the story, and communication of that story to the public (ibid.). In data journalism, data is the source, and computational methods and applications are the tools to aid journalists in their work (ibid).

The accessibility of data, the availability of information and computational and data analytics techniques, and the wide array of tools available to journalists for the easy manipulation and publication of data has had a significant role in the journalistic use of data, information, and computer science techniques in journalistic reporting in recent times (Lewis and Westlund 2014). The recent take-up of data-driven journalism can be seen in the recent formation of data teams at many leading news organisations; the development of data centric journalism programmes and courses (Heravi 2018); and the emergence of news graphics, interactives, applications, and games (Lewis and Westlund 2014; Usher 2016).

The recent interest in, and uptake of, data journalism in newsrooms across the world and its more recent emergence of data journalism and related fields as an academic discipline calls for a systematic study of the domain, particularly in terms of the state of the art and current practices in this area in newsrooms across the world, and an in-depth understanding of the skills, requirements, and training in this area. To address this need, the Global Data Journalism survey studied the status of this specific field and the orientation and know-how of practitioners.

The research questions addressed in this paper are as follows:

RQ1: What are the newsroom practices when it comes to the use of data for journalistic purposes?

RQ2: What are the educational background and skills of the journalists engaged, or interested in engaging, in data-driven journalism practices?

RQ3: Which data skills are most important for data journalists to acquire for future progression?

RQ4: What is the perceived impact of data journalism on journalistic values?

A brief summary of the results of this study was previously reported in a short paper presented at iConference (Heravi 2018). Extending that paper, the paper at hand provides a detailed account of this survey, reports on fuller arrays of results, and provides further reflections and discussion on the matters. Additionally, the anonymised dataset behind this study is made available to the public as an online appendix to this paper 1.

STUDIES ON DATA JOURNALISM

The term data journalism is a relatively new term, with fewer than 10 years of history, and an even shorter history in academic literature. In an analysis of data journalism literature, Ausserhofer et al. (2017) suggest an increase in research publications on data journalism and related fields since 2010. They report that only a small number of isolated research publications on data journalism and related areas were found before 2010, most of which were authored by American journalism researchers, investigating the use of various computational technologies in the newsroom (e.g., Davenport et al. 2000; Davenport et al.1996; Garrison 1999). Their study further suggests that, even though CAR has been practiced since the 1960s, the scientific investigation of it has started only recently (Ausserhofer et al. 2017).

Since 2013, there has been a growing body of studies on the practice of data journalism in various countries or specific geographical areas. Examples are Sweden (Appelgren and Nygren 2014), Norway (Karlsen and Stavelin 2014), Belgium (De Maeyer et al. 2015), Canada (Young et al. 2018), Russia (Radchenko and Sakoyan 2014), the United Kingdom (Arias-Robles and López 2020; Borges-Rey 2017; Borges-Rey 2016; Hannaford 2015; Knight 2015; Dick

2013), the United States (Fink and Anderson 2014; S. Parasie and Dagiral 2013; Parasie 2015), Germany (Stalph 2017; Weinacht and Spiller 2014), Italy (Porlezza and Splendore 2019), Australia (Wright and Doyle 2019), Latin America (Mutsvairo et al. 2020; Borges-Rey 2019;Palomo et al. 2019; Borges-Rey et al. 2018), China (Zhang and Feng 2019), the Arab region (Fahmy and Attia 2020; Lewis and Nashmi 2019), Pakistan (Jamil 2019), and further studies beyond the Majority World countries (Appelgren et al. 2019; Wright et al. 2019).

Apart from the peer-reviewed studies of data journalism practices in specific countries or regions, Google News Lab, in parallel with our study2, conducted a study on data journalism in news organisations in the United States, the United Kingdom, France, and Germany (Rogers et al. 2017). Google's survey was focused on users of Google products, and was initially planned as an internal study for Google itself, according to Rogers. The report was, however, later published and shared publicly. They found that their respondents felt that data impacts journalism in three ways: (1) "It helps reduce complexity and give readers a chance to make sense of the world around them", (2) it "Keeps society rooted in facts", and (3) it "Improves reputations of newsrooms with advertisers, increasing revenue potential with innovative data journalism and visualization" (Rogers et al. 2017, p. 10). They further found that a majority of their respondents in these four countries would like their organisations to make more frequent and effective use of data in their storytelling, while they identified a lack of skills in relevant areas as an important barrier to using data in their organisations (Rogers et al. 2017).

In addition to the above studies, there have been a smaller number of recent studies examining

the characteristics of good or award-winning data journalism (Young et al. 2018; Ojo and Heravi 2018; Loosen et al. 2017). Amongst these, Ojo and Heravi (2018) examined the evolving skill set and competencies required for data journalism and their adoption pattern by winners of the Data Journalism Awards, and identified a set of core skills, technologies, and tools that appear central to award-winning data journalism practices.

In terms of research methods, the literature review by Ausserhofer et al. (2017) reports qualitative interviews as the most common methodological framework for data journalism research, followed by content analysis. Newsroom observations and ethnographic methods were also seen in the literature review. However, in a significant number of publications the interviews or the content analysis are limited to only a small number of participants or publications, and in many cases are limited to only one organisation or publication. Examples are Royal's (2012) study of the New York Times, Parasie and Dagiral's (2013) study of Chicago news institutions, Hannaford's (2015) study of the BBC and The Financial Times, Parasie's (2015) study of the Centre for Investigative Reporting, Hullman et al.'s (2015) study of The Economist, Young and Hermida's (2015)

study of Los Angeles Times, Tandoc and Oh's (2015) study of the Guardian's stories, and Palomo et al.'s (2019) study of La Nacion.

The higher prevalence of qualitative methods in data journalism research highlights the lack of quantitative approaches, such as surveys and questionnaires, in data journalism research. Ausserhofer et al. (2017) report that the method of survey was in fact one of the least commonly used methods between the publications they studied, and where present the surveys were used to support the qualitative research that was primarily performed (e.g., Appelgren and Nygren 2014).

Another concern about the existing research in this area is the tendency to select larger, better known publications/organisations, such as the Guardian, the New York Times, the Los Angeles Times, The Economist, The Financial Times, and the BBC (Borges-Rey 2016; Hannaford 2015; Hullman et al. 2015; Knight 2015; S. Parasie and Dagiral 2013; Royal 2012; Tandoc and Oh 2015; Young and Hermida 2015). However, it remains unclear if these organisations were selected more frequently because they are engaged in producing a higher volume of data journalism projects (and hence have more visibility) or simply because of their reputation, which itself could be the result of many other factors. Hence, our knowledge of the domain is mostly limited to the more known players, as opposed to a wider and more generalised view of the domain.

While there has been an increasing body of academic research on data journalism and related fields in the past five years, most are qualitative studies with a narrow focus. These studies in large part have examined specific and, in most cases, small numbers of (often well known) news organisations or data journalists in their respective countries. Many of the publications on data journalism, particularly those before 2016, do not present any mention of theoretical frameworks, which may suggest that data journalism research has been more geared towards practice, and less on theoretical and academic research (Ausserhofer et al. 2017).

Overall, there is a lack of comparative studies on the practise of data journalism across national borders and famous organisations. The Global Data Journalism survey aimed at an independent, global, and inclusive study of data journalism practices across the globe, with the aim of better understanding the emerging area and as way to provide theoretical, practical, and pedagogical guidelines for the future of data journalism. A brief summary of the results of this study was published in 2018 in the form of a short paper (Heravi 2018). The paper in hand provides a considerably extended view of the results of this study. Furthermore, the full, anonymised data behind this study will be made available to the public as an appendix to this paper.

METHOD

The Global Data Journalism survey was launched on 3 December 2016 and closed on 10 May 2017. While the survey was open to all data journalists and journalists globally, it was limited to those who identified as having worked as a journalist or a data journalist in the year prior to the completion of the survey.

The survey was an online (internet/web-based) survey and was carried out using Google Forms. The Global Data Journalism survey aimed to attract journalists from across the world, yet there is no single governing body or a list that includes all potential participants across the world. Given the open nature of the Global Data Journalism survey, targeting individuals from specific countries, PR databases, or private or membership organisations (such as the National Union of Journalists – NUJ or The National Institute for Computer-Assisted Reporting – NICAR) was considered an inappropriate method for this study. On the contrary, the open and "as broadly as possible" approach to the distribution of this survey was adopted in the effort to reach the widest and most heterogeneous group of potential respondents.

The survey was circulated and promoted as broadly as possible through various platforms and channels. A link to the survey was distributed widely through social media channels using #ddj hashtag (the most active hashtag used by the data journalism community worldwide), relevant listservs, and two Slack groups—News Nerdy and DJA (Data Journalism Awards) 2017. A number of articles about the survey were featured in the media (Lorenz 2016; Plaum 2016; SiliconRepublic 2016), in addition to direct targeting by the Global Editors Network and Data Journalism Awards, who sent an email on the survey to their subscribers (Bouchart 2016). Through their Global Data Journalism Awards mailing list, the Global Editors Network have access to a large, global audience of data journalists across the world. This mailing list, combined with the #ddj hashtag, are considered to be the widest reaching mediums in terms of the global data journalism community.

The survey consisted of 48 questions in 7 sections. Two hundred and six participants from 43 countries participated in this survey, with 181 respondents filling it out to completion. For the purpose of analysing the results, only responses completed to the end were considered and the rest were discarded; hereon in, when we refer to "participants", we are only referring to the 181 who completed the survey.

RESULTS

Respondents' Profile

The results of the survey suggest that half of the participants (51%) were from the continent of Europe and a third (33%) were from North America, leaving only 16% from the rest of the world, including Central and South America, Africa, Asia, and Oceania. In terms of specific countries, the United States had the highest number of participants, at 31% of the total respondents. The United States was followed by the United Kingdom, with 9% of the participants. These were followed by Germany and Spain (both 6%), and then Ireland and Italy (both 4%).

Comparing these figures with the originating countries of the winning cases of the Data Journalism Awards (Ojo and Heravi 2018) provides an interesting insight. There were 44 winning cases at the Data Journalism Awards between 2013 and 2016, originating from 14 different countries. In that study, the United States dominated 46% of the Global Editors Network (GEN) Data Journalism awardees between 2013 and 2016, followed by the United Kingdom with 12% of the awards. The figures for these countries in terms of their participation in the data journalism study and winning data journalism awards are almost identical and highly correlated. These two countries present the highest quality of data journalism work, likely the highest number of active data journalists, and the most willingness to participate in data journalism-related studies. France and Argentina, however, were both the third contenders in terms of winning Data Journalism Awards between 2013 and 2016, but their rate of participation in the study did not match the rate of their winning cases, as it did for the United States and the United Kingdom. This may be due to language barriers, the small group of journalists who may be producing the award-winning journalism output, or a lack of interest in participation in such studies from journalists in these countries for various reasons unknown to us.

In addition to the observed similarities between the participation rate and awardees of data journalism awards, our country-specific participation figures show similarities with the number of published studies on data journalism from various countries. Ausserhofer et al. (2017) report that most of the published research on data journalism tends to come from the United States or the United Kingdom. They further report a few published works from Western European countries such as Germany and Sweden. They, too, specifically point out the lack of published research from a country such as France, noting "the fact that even large Western [European] countries such as France are not represented in our sample might be partly due to the lack of a longstanding tradition of CAR there" (Parasie and Dagiral 2013; Ausserhofer et al. 2017, p. 15).

In terms of their gender and age, 57.5% of the participants identified as male and 42.5% as female. Seventy-seven per cent (77%) of participants were aged 25 to 44 years old, which shows a professional but rather young corpus of participants. Only 7% were younger than 24 and 16% were over 45.

Of all the participants, 64% were in full-time employment, 18% were freelance, 12% were part-time, and 4% were casual/retainer. In terms of the size

of organisation, 32% worked in large organisations of 500+ employees, 22% in organisations of size 10–49, 17% in organisations with 100 to 499 employees, 15% in small organisations of 2 to 9 employees, and only 8% in mid-sized organisations of 50–99 employees. Just under half of the participants (42%) reported to work in national organisations, while 20% worked in local, 18% in international, and the rest in a combination of these types or other types of organisations.

In terms of experience as a journalist, the majority of our respondents (78%) were individuals with 1 to 10 years of experience as a journalist, with a breakdown of 2% having less than a year of experience,41% having 1 to 4 years of experience, and 26% having 5 to 9 years. A total of 19% of our participants have 10 to 19 years of experience and only 11% have over 20 years of experience as a journalist.

In terms of beat, politics appeared to be the most prevalent area to cover between respondents, with 71% identifying it as one of their primary areas to cover. This was followed by business (36%) and world news (35%).

In terms of publication medium, 43% of the participating journalists produced content for online platforms of broadcast or print media outlets, and 34% produced content for online-only publications. This makes a total 77% of all participants producing content for online publications. This figure was followed by print newspaper (8%), radio (4%), TV (4%), print magazines (3%), and personal blog (2%), with producing content for a news agency making up only 1% of the total.

Newsroom Practices

To study newsroom practices and challenges when it comes to the use of data in investigations and reporting (RQ1), we asked our participants about the status of data journalism in their organisations. Forty-six per cent (46%) claimed that they have a dedicated data desk/team/unit/blog/section. This figure is followed by 29% who expressed that they do not have such a dedicated group but publish data-driven projects on a regular basis. This means that, regardless of having a dedicated team or section, 75% were from organisations that publish data-driven stories. Seven per cent (7%) of the participants noted that they plan to work with data in the next six months, and 7% expressed that they have no immediate plan to start working with data (Figure 1).

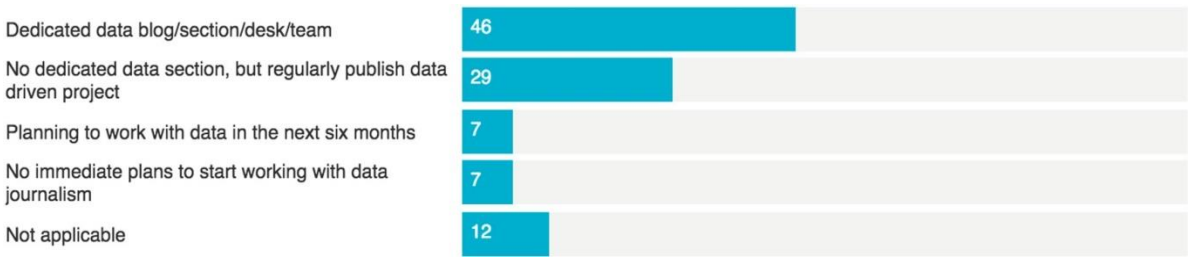

Figure 1. How data journalism is performed in news organisations (%), Global Data Journalism Survey, N = 181.

Of those who stated to have dedicated data desk/team/unit/blog/section, 40% had a data team consisting of 3 to 5 people and 30% had a team of 1 to 2 people. This means that the vast majority (70%) of organisations with data teams operate with small teams of 1 to 5. On the other side of the spectrum, 22% of participating organisations had data teams of 6 to 10 people, 3% had a team of 11 to 15 people, and 5% had large data teams of more than 15 people.

Concerning areas of coverage that could benefit best from data journalism, politics came to the top, with 81% expressing it as an area that could benefit the most from using data in investigations and storytelling. This is followed by economy and crime, both with 78%; education and health, both with 77%; national topics (64%); and local topics (62%) that would benefit tremendously from data journalism.

To study the challenges in employing data journalism, we asked our participants about the main hurdles in implementing data journalism in their organisation. Over half of the participants (52%) considered "lack of resources" as the main barrier in implementing data journalism in their organisations. This was followed by "lack of adequate knowledge" (46%) in terms of tools and working procedures and "lack of time" (40%). "Lack of adequate publishing infrastructure" that is fit for purpose to support publication of data driven stories was surprisingly flagged by 36% as one of the hurdles in implementing data journalism. Some 25% believed that data journalism is not implemented in their organisation, partially due to "lack of interest from staff", and another 23% expressed "lack of management support" as one of the barriers in the successful implementation of data journalism intheir organisation (Figure 2). These figures show that while lack of management support and potential lack of interest from the editorial staff could significantly affect the implementation of data journalism in news organisations, the skill and resource-related issues have a larger impact on the successful implementation of data journalism in newsrooms.

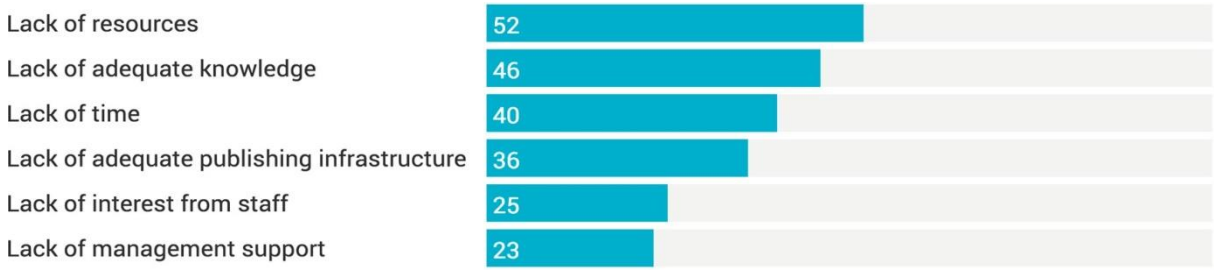

Figure 2. Main hurdles in implementing data journalism (%), Global Data Journalism Survey, N = 181.

One participant expressed that "not all organizations are interested or value data journalism and it has hampered my experience in journalism recently. Data in journalism is incredibly important, but when the higher-ups don't see it as valuable (or don't understand it) it's difficult to integrate it into reporting" (participant, 10–49 organisation, US).

One area of tension when it comes to the implementation of data journalism in newsrooms is the attribution of bylines. We asked our participants: "Do you get your name in the byline when you have done the data analysis, i.e., finding the story in the data, but have not written the words? Or only the person who writes the words gets the byline?" Out of those who identified with this scenario, 60% expressed that their name will be included in the byline if they were involved in the data analysis and finding the story in the data but not in writing the words. The other 40%, however, expressed otherwise. Of the latter 40% who did not get their names in the byline, 21% expressed that while they do not get their name included in the byline, it will be mentioned somewhere in the article, while the other 19% got no acknowledgement of their work in finding the story in the data.

One participant from a large Canadian news organisation with a dedicated data team identified byline allocation as a "frequent cause of tension" in their news organisation. They expressed "I've had entire data-driven stories taken and put under a regular reporter's name close to publication. [There] seems to be a mistrust of the findings, like if you get such a good, strong story, it can only be told by someone who would normally get such a story through a leak" (participant, 500+ organisation size, Canada). They added that, as reporters and editors are getting increasingly educated in this area, such circumstances are happening less frequently and the situation is overall improving (paraphrased, participant, 500+ organisation size, Canada).

Education and Skills

To explore RQ2 and RQ3, we asked the participants about their educational background, existing skills, and skills they believe to be important to acquire for their future work.

Eighty-six per cent (86%) of the participating journalists indicated that they consider themselves to be data journalists. Despite this high proportion, in terms of data journalism proficiency only 18% rate themselves as experts, while 44% of the respondents identify as having a better than average knowledge and 26% identify as having average knowledge in data journalism. Thirteen per cent (13%) of the participants identified as a novice or having a below-average level of expertise (Figure 3).

Half of our participants (50%) had formal training in data journalism and the other half did not. In terms of a wider understanding of formal training in the knowledge areas used in data journalism, most participants demonstrated a high

degree of formal training in journalism, with lower and varying degrees of formal training in the more data-oriented and technical aspects, such as data analysis, statistics, coding, data science, machine learning, and data visualisation. Figure 4 depicts the breakdown of formal training in various related fields between our participants.

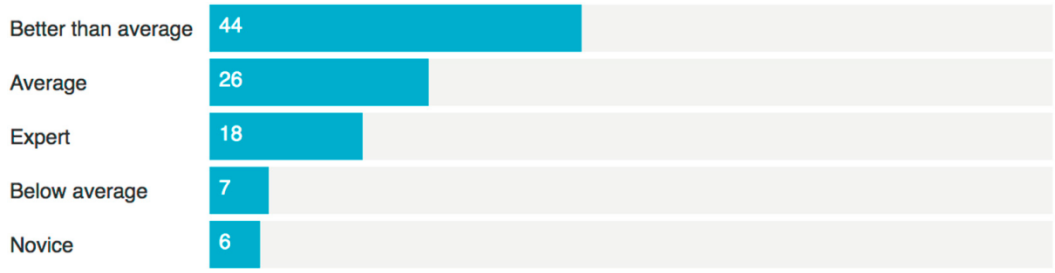

Figure 3. Self-assessed knowledge of data journalism (%), Global Data Journalism Survey, N = 181.

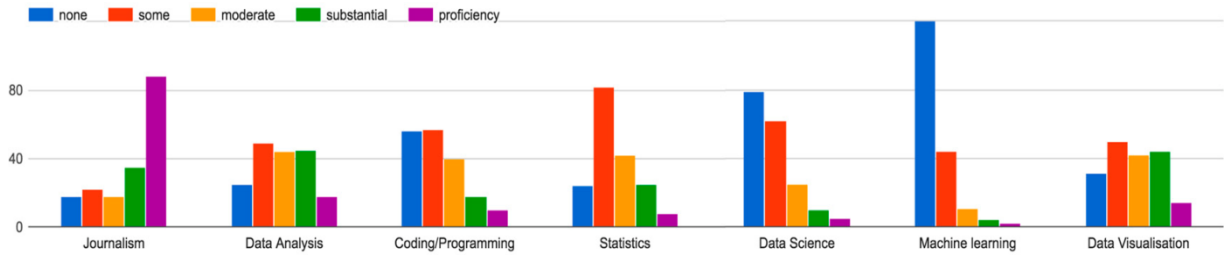

Figure 4. Level of formal training in related knowledge fields (%), N = 181.

In terms of general education level, 96% of our respondents had a university degree, with a breakdown of 40% at undergraduate (bachelor) level, 53% at postgraduate level, and 3% with a doctorate or above degree. This shows that the data journalism community is a highly educated community. Looking into the degrees obtained by these participants, 62% were formally educated in journalism at the university level. While journalism was by far the most prevalent obtained higher education degree between our participants, it was followed by a combination of other degrees, such as politics (15%), computer/information/data science/engineering (12%), and communication and language/literature each with 10.5%, with 26% listing a combination of other degrees. This indicates that while most participating journalists have formal higher education training in journalism, communication, politics, and related degrees such as literature, only 12% have higher education training in the more data-related and technical topics. This further reflects on the basic underlying reasons behind the level of training demonstrated in Figure 4. It denotes that formal training amongst the participants seems to have been mainly obtained through higher education and university

degrees, and highlights the importance of including data-related courses and modules in relevant higher education journalism and communication programmes.

An experienced participant (20+ year) from a mid-size organisation with a dedicated data team noted "I think some of the data folks need more journalism background/training, as well as the other way around. I think students also need more rigor in finding authoritative sources of information and data, i.e., research skills" (participant, 10–49 organisation size, USA). Another participant from a large UK-based organisation with a dedicated data team highlighted "the need to teach journalism as much as data rather than as separate disciplines—which is how many newsroom are treating the subject" (participant, 500+ organisation size, UK).

Despite the growing number of programmes adding data journalism to their curricula, so far there are a limited number of complete short or long-term training/educational programmes dedicated to the specific field of data journalism (Heravi 2018). One reason behind this may be that areas encompassing the disciplines involved in this multidisciplinary domain are rather difficult to find under one singleroof. There is no consensus yet as to the ideal curriculum for data journalism. Two sets of suggestions for various ways to approach teaching data and computational journalism to a broader group of individuals are proposed by Berret and Phillips (2016) and Heravi (2017). Heravi (2017) suggests that data journalism training is often geared towards journalists, where they are taught data. However, other complementary approaches could be envisaged, where journalistic practices are taught to those with data or computer skills, or mixed programmes including both skillsets from the beginning (ibid.).

To explore the appetite and willingness of journalists and news organisations to learn new data and computational skills, we asked whether or not they would be interested in gaining further skills in relation to data journalism, and if so which specific skills they would be most interested in acquiring. A striking majority of the participants in the survey (98%) expressed that they were interested in acquiring further skills to practice data journalism, with 81% indicating that they are *very* interested. Despite this near unanimous interest, only 42% expressed that they are interested in formal higher education degrees in this area. However, if the training offered is shorter-term or more flexible, a significant 74% of the participating journalists would be interested in formal training in higher education—e.g., a postgraduate certificate or higher education diploma.

In terms of the specific data skills journalists are interested in acquiring, RQ3, data analysis presented itself as the top skill, with 64% of individuals expressing interest in learning about it. This was marginally followed by learning "how to programme/code" at 63%, and visualising data at 51%. These top three data skills were followed by another three skills: "how to clean data", "how to develop data-driven applications", and to learn "how to check if data is reliable", with over 48% of journalists expressing interest in each (Figure 5).

In summary, these results show that most participating journalists had formal higher education training in journalism and related areas, but these same journalists lacked training in data skills. Shorter, targeted higher education programmes would be the most attractive offering for increasing their skills in data, and many expressed a desire to make such improvements. The most important topics to be taught, according to the survey, are data analytics skills, followed by coding skills. These figures signify how important training—and particularly data journalism training—is, particularly when it comes to formal higher education training.

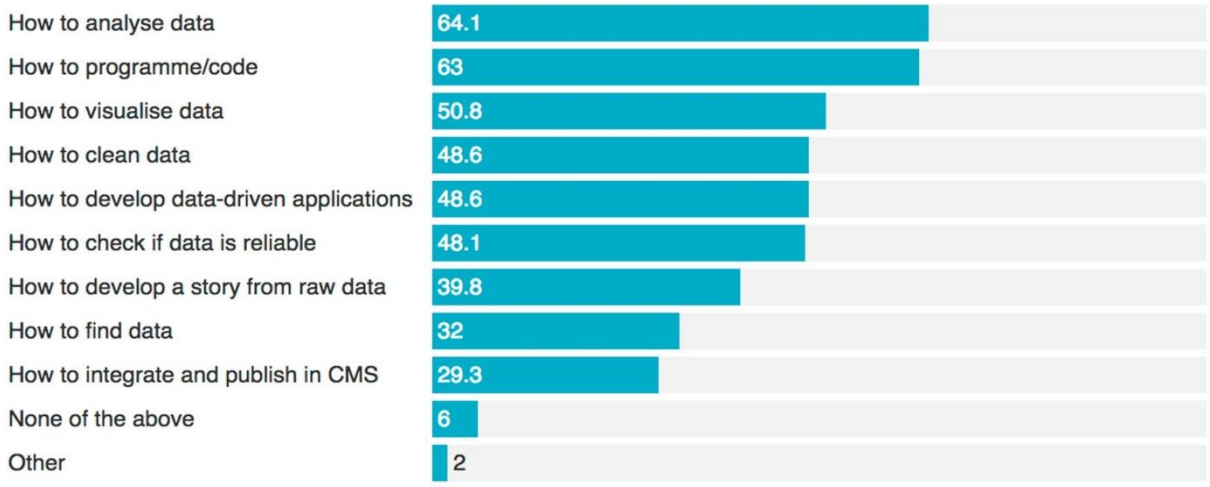

Figure 5. Interest in acquiring the listed skills (%), Global Data Journalism Survey, N = 181.

Data as a Source and Journalistic Values

To study the role of data as a source in journalistic practices (RQ4), we asked our participants about the importance of the use of raw data as a primary source in journalism. Nearly 95% of the participating journalists expressed that they believed "the use of raw data in journalism" is veryimportant or important. A majority (70%) expressed that they will not be able to carry out their work without data as a source (Figure 6).

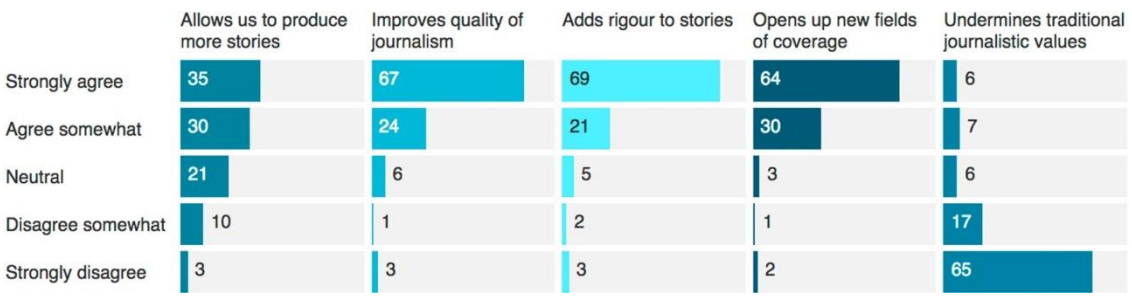

Figure 6. Data journalism, quantity, quality, rigour, opportunities and values (%), Global Data Journalism Survey, N = 181.

To examine the perceived values associated with the use of data in journalistic practices (RQ4), we asked our participants a series of questions, a number of which are discussed in this section. Sixty-five per cent (65%) of the respondents strongly agreed or somewhat agreed that data journalism allows them or their organisation to produce more stories. On the other end of the spectrum, 13% somewhat disagreed (10%) or strongly disagreed (3%) with this statement.

In terms of the quality associated with journalistic output, 90% of the respondents strongly agreed (69%) or agreed somewhat (21%) that data journalism adds rigour to journalism, with only 5% expressing the opposite. Similarly, 91% strongly agreed or somewhat agreed that data journalism improves the quality of journalistic work in their organisation, with only 4% believing the opposite (Figure 6).

One respondent said, "I think [use of data in journalistic process] will make me a stronger journalist" (participant, freelance, US). Another participant from a large news organisation expressed the belief that the use of data in journalistic processes "should be the foundation of all investigations" (participant, 500+ organisation, Ireland).

Fifty-two per cent (52%) of the participants strongly agreed or somewhat agreed that data journalism opens up new fields of coverage for them, and nearly 70% believed that data journalism could be of such quality to generate extra sales or income for their news organisations.

Tapping into traditional journalistic values while leaving the definition of these values to the participants, 83% of the participating journalists disagreed somewhat or strongly disagreed that data journalism undermines traditional journalistic values, while only 11% agreed somewhat or strongly agreed that data journalism is undermining these values (Figure 6).

DISCUSSION

The results from this study indicate a strong uptake of data journalism as an innovative practice, and the acceptance of such practices in news organisations around the world. Examining newsroom practices in relation to the use of data for journalistic purposes (RQ1), the study reveals that a growing number of news organisations have hired data journalists and formed data journalism teams, in most cases with two to five members. While this is a positive change in comparison to a few years ago, it is a slow development and it should be critically questioned whether such small team sizes are strong enough to fully support the consistent use of data in journalistic practices. While small and medium-sized data journalism teams appear to be able to produce high quality data stories, research

suggests that larger newsrooms are able to produce higher quality data stories, simply by hiring a larger number of data journalists, data scientists, and developers (Fink and Anderson 2014; S. Parasie and Dagiral 2013). The purpose of data journalism teams is worth considering if the resources remain small; are the teams meant to focus on long investigations resulting in a small number of stories, or are they there to build the use of data across everyday workflows?

While data journalism has reached a level of maturity, the role of data journalists in the news production cycle has to be more clearly defined. The tension with bylines, for example, suggests that the role of data journalists is not clear in some news organisations, or they are not fully embraced as "journalist" or "editorial" staff. Effective collaboration amongst the data team or data journalists and the rest of the editorial team could make a significant difference in terms of quality, as well as quantity, of stories produced in a news organisations. In order for these teams to work more effectively, they may require better integration with the editorial team, and in general within the news productions cycle. Expressing their concerns, a participant from a large new organisation with a dedicated data team in Denmark noted that they think an important factor which separates data desks from each other is whether the data desk can work independently [in an editorial capacity] or has to work as a service desk. The participant added that many organisations believe that data-research and visualisation is a service for journalists in the news organisation, and expressed the belief that in order to do excellent and original data journalism work, a team needs to work with the story from the beginning to the end—i.e., from the idea phase to visualisation and writing (participant, 500+ size organisation, Denmark). The idea that the best work emerges when data journalists or data teams are integrated in the newsroom as editorial staff (instead of as a service desk) is corroborated by studies on the characteristics of high quality data journalism (e.g., Ojo and Heravi 2018; Young et al. 2018; Loosen et al. 2017).

To investigate skills and educational background associated with this area of practice (RQ2), we examined the educational background of participating journalists, the skills they report to already possess, and the skills they consider important to acquire for their future work (RQ3). The study shows that the data journalism community is a highly educated community. This community has its roots mostly in journalism and communication degrees, and less so in data/information and computer-related disciplines. While the results show that most participating journalists have formal higher education training in journalism, and even though a striking majority describe themselves as "data journalists", the study reveals a lack of systematic training in data skills amongst these journalists. At the same time, the participants exhibit a considerable interest in acquiring skills in data analysis, statistics, coding, and data visualisation, and were particularly interested in shorter-term formal higher education training in this area.

Many respondents named "coding skills" as desirable skills for further education. The perception that "learning to code" is necessary for practicing data journalism and can be seen as a big hurdle in entering the field. The discussion around whether or not coding is a necessary skill in data journalism is not a new one, and is not one that could have one single answer. Putting the value of coding in data journalism, next to how it may deter new entrants to the field, as well as our experience and observations in teaching data journalism in academic and professional settings, we believe that while coding skills can be of high value in the production of data driven stories, not all data journalists need to be able to code. This is not to say data journalists should not learn coding. Rather, this is to suggest they should not see "coding" as a big hurdle for entering this field. It is important to remember that data journalism is not equal to "computer-science-doing-journalism". As stated in our definition of data journalism, data journalism puts the tenets of journalism first. In line with the results from the survey, which puts journalism first and other data and computational skills next, Heravi (2017) suggests that the first skills that any student must learn are journalistic and investigative skills. After journalism skills, the most important topics to cover would be familiarity with data and data sources, an understanding of the lifecycle of data journalism projects, skills for data wrangling, and most importantly data analysis skills, including sufficient knowledge of statistics. More advanced data visualisation, programming, and other advanced topics may follow. Having said that, embedding more advanced computational techniques in production workflows allows for finding stories not everyone can find and the presentation of stories in a more accessible and appealing manner.

While technical, data analytics, and statistical skills do not appear to be the strength of participating journalists when compared with their journalism skills, it appears that many newsrooms already have a dedicated data team and produce data-driven stories on a regular basis.

Studying the perceived impact of data journalism on journalistic values (RQ4), the results suggest a shared belief is that data journalism has brought opportunities for newsrooms and journalists. The study further reveals that despite debates in the use of data for producing journalistic work, both in terms of quantity and quality, a vast majority of journalists believe that data journalism allows them to create more stories in terms of quantity, which are also more rigorous and of higher quality.

In terms of threats to traditional journalistic values, a vast majority of the participants believe that data journalism adds rigour to their stories and disagree with the idea that suggests data journalism may undermines traditional journalistic values.

CONCLUSIONS

Providing a descriptive account of the results of the Global Data Journalism Survey, this paper helps to re-evaluate the current state of data journalism globally. The results specifically show awareness of, interest in, and uptake of the use of data in newsrooms around the world, and an agreement not only that data provides new opportunities for newsrooms, but also that it adds rigour and improves the quality of journalistic output.

While this is a transitional time for data journalism, the practices have matured enough that the primary question in many newsrooms may no longer be about whether or not there is value in adding data journalism to newsrooms. On the contrary, what is needed is a greater recognition and clarification of the role of data journalists within the news production cycle, including their function, level of editorial input, and appropriate recognition for their contributions. This paper suggests that, for data journalism to flourish, it must be seen as a key methodological approach to contemporary journalism, as opposed to a service role supporting the editorial team. At the same time, for these practices to mature further, we need a higher level of support from management in terms of resources and the time allocated to exploratory projects, as well as a higher level of training and skill development for journalists.

Further inferential analysis of the results of this survey would be of great interest. The researchers have made the data behind this study publicly available to all researchers and interested parties, which will enable the further analysis of this data by other researchers in the future. Further study of the journalistic values when it comes to the use of data as a source as well as a means of communication in journalism is another interesting area of future research, which would provide valuable insight into the field of data journalism research and practice.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Appelgren, Ester, and Gunnar Nygren. 2014. Data journalism in Sweden. Digital Journalism 2: 394–405. [CrossRef] Appelgren, Ester, Carl-Gustav Lindén, and Arjen van Dalen. 2019. Data journalism research: Studying a maturing field across journalistic cultures, media markets and political environments. Digital Journalism 7: 1191–99. [CrossRef]

Arias-Robles, Félix, and Pilar José López López. 2020. Driving the closest information. Local data journalism in the UK. Journalism. [CrossRef]

Ausserhofer, Julian, Robert Gutounig, Michael Oppermann, Sarah Matiasek, and Eva Goldgruber. 2017. The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive. Journalism. [CrossRef]

Berret, Charles, and Cheryl Phillips. 2016. Teaching Data and Computational Journalism. New York: Columbia Journalism School, ISBN 978-0-692-63745-6. Available online:

journalism.columbia.edu/system/files/content/
teaching_data_and_computational_journalism.pdf (accessed on 19 October 2020).
Borges-Rey, Eddy. 2016. Unravelling data journalism. Journalism Practice 10: 833–43. [CrossRef]
Borges-Rey, E. 2017. Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. Journalism 21: 915–32. [CrossRef]
Borges-Rey, Eddy. 2019. Data journalism in Latin America: Community, development and contestation. In Data Journalism in the Global South. Palgrave Studies in Journalism and the Global South. Edited by Bruce Mutsvairo, Saba Bebawi and Eddy Borges-Rey. Cham: Palgrave Macmillan. [CrossRef]
Borges-Rey, Eddy, Bahareh Heravi, and Turo Uskali, eds. 2018. Ibero-American data journalism: development, contestation, and social change. ICONO14. [CrossRef]
Bouchart, Marianne. 2016. GEN and the DJA support #ddj global survey. Global Editors Network, Data Journalism Awards. Available online: https://us3.campaign-archive.com/?u=9163cb87f87c5fb73d398ba6e& id=decdb46f40&e=f1cb0b584b (accessed on 19 October 2020).
Coddington, Mark. 2015. Clarifying Journalism's quantitative turn. Digital Journalism 3: 331–48. [CrossRef] Davenport, Lucinda, Fred Fico, and David Weinstock. 1996. Computers in Newsrooms of Michigan's Newspapers.
Newspaper Research Journal 17: 14–28. [CrossRef]
Davenport, Lucinda, Fred Fico, and Mary Detwiler. 2000. Computer–assisted reporting in Michigan daily newspapers: More than a decade of adoption. In Newspaper Division, AEJMC National Convention. Phoenix: AZ. Available online: http://online.sfsu.edu/jjohnson/Courses&Syllabi/BU-JO807/Bibli&Articles/Davenport(2000).pdf (accessed on 19 October 2020).
De Maeyer, Juliette, Manon Libert, David Domingo, Francois 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: 432–46. [CrossRef]
Dick, Murray. 2013. Interactive infographics and news values. Digital Journalism 2: 490–506. [CrossRef]
Fahmy, Nagwa, and Maha Abdul majeed Attia. 2020. A field study of Arab data journalism practices in the digital era. Journalism Practice. [CrossRef]
Fink, Katherine, and Christopher Anderson. 2014. Data journalism in the United States. Journalism Studies 16: 467–81. [CrossRef]
Flew, Terry, Christina Spurgeon, Anna Daniel, and Adam Swift. 2012. The promise of computational journalism. Journalism Practice 6: 157–71. [CrossRef]
Garrison, Bruce. 1999. Newspaper size as factor in use of computers for newsgathering. Newspaper Research Journal 20: 72–85. [CrossRef]
Hannaford, Liz. 2015. Computational Journalism in the UK newsroom: hybrids or specialists? Journalism Education 4: 6–21.
Heravi, Bahareh. 2017. Teaching Data Journalism. In Data Journalism: Past, Present and Future. Edited by John Mair, Richard Lance Keeble, Megan Lucero and Moore Moore. Martin: Abramis Academic Publishing.
Heravi, Bahareh. 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 J., Val Gillet and Peter Willett. Cham: Springer, vol. 10766. [CrossRef] Heravi, Bahareh. 2018. 3Ws of Data Journalism Education: What, where and who? Journalism Practice 13: 349–66. [CrossRef]
Howard, Alexander Benjamin. 2014. The Art and Science of Data-Driven Journalism. New York: Tow Centre for Digital Journalism, Columbia University. [CrossRef]
Hullman, Jessica, Nicholas Diakopoulos, Elaheh Momeni, and Eytan Adar. 2015. Content, context, and critique: Commenting on a data visualization blog. Paper presented at the 18th ACM

Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15), Vancouver, BC, Canada, March 14–15; New York: ACM, pp. 1170–75. [CrossRef]

Jamil, Sadia. 2019. Increasing accountability using data journalism: Challenges for the Pakistani journalists. Journalism Practice. [CrossRef]

Karlsen, Joakim, and Eirik Stavelin. 2014. Computational Journalism in Norwegian Newsrooms. Journalism Practice 8: 34–48. [CrossRef]

Knight, Megan. 2015. Data journalism in the UK: A preliminary analysis of form and content. Journal of Media Practice 16: 55–72. [CrossRef]

Lewis, Norman, and Eisa Al Nashmi. 2019. Data journalism in the Arab Region: Role conflict exposed. Digital Journalism 7: 1200–14. [CrossRef]

Lewis, Seth C., and Oscar Westlund. 2014. Big Data and Journalism: Epistemology, expertise, economics, and ethics. Digital Journalism 3: 447–66. [CrossRef]

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. [CrossRef] Lorenz, Mirko. 2016. What is the current status of data journalism? Participate in our global survey to find out. Datadrivenjournalism.net. This article is no longer available online.

Mutsvairo, Bruce, Saba Bebawi, and Eddy Borges-Rey, eds. 2020. Data Journalism in the Global South. London: Palgrave Macmillan. [CrossRef]

Ojo, Adegboyega, and Bahareh Heravi. 2018. Patterns in Award Winning Data Storytelling: Story Types, Enabling Tools and Competences. Digital Journalism 6: 693–718. [CrossRef]

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: 1270–88. [CrossRef]

Parasie, Sylvain. 2015. Data-driven Revelation? Digital Journalism 3: 364–80. [CrossRef]

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: 853–71. [CrossRef]

Plaum, Alexander. 2016. Call for Participation: Global Data Journalism Survey 2017. Deutsche Welle Innovation Blog. Available online: http://blogs.dw.com/innovation/call-for-participation-global-data-journalism-survey- 2017 (accessed on 19 October 2020).

Porlezza, Colin, and Sergio Splendore. 2019. From open journalism to closed data: data journalism in Italy. Digital Journalism 7: 1230–52. [CrossRef]

Radchenko, Irina, and Anna Sakoyan. 2014. The view on open data and data journalism: Cases, educational resources and current trends BT—Analysis of images, social networks and texts. In International Conference on Analysis of Images, Social Networks and Texts. Edited by Dmitry I. Ignatov, Mikhail Yu Khachay, Alexander Panchenko, Natalia Konstantinova and Rostislav E. Yavorsky. Cham: Springer International Publishing, pp. 47–54.

Rogers, Simon, Jonathan Schwabish, and Danielle Bowers. 2017. Data Journalism in 2017: The Current State and Challenges Facing the Field Today. Available online: https://newslab.withgoogle.com/assets/docs/data- journalism-in-2017.pdf (accessed on 19 October 2020).

Royal, Cindy. 2012. The journalist as programmer: a case study of the New York Times interactive news technology department. #ISOJ: The Official Research Journal of the International Symposium for Online Journalism 2: 5–24.

SiliconRepublic. 2016. Q&A: How Is Data Journalism Changing the Newsroom? An Interview with Bahareh Heravi. Edited by Elaine Burke. Dublin: Silicon Republic. Available online: https://www.siliconrepublic.com/ innovation/data-journalism-media-newsroom (accessed on 19 October 2020).

Stalph, Florian. 2017. Classifying Data Journalism: A content analysis of daily data-driven stories. Journalism Practice, 1–19. [CrossRef]

Tandoc, Edson C., and Soo-Kwang Oh. 2015. Small Departures, Big Continuities? Journalism Studies 18: 997–1015. [CrossRef]

Usher, Nikki. 2016. Interactives Journalism: Hackers, Data, and Code. Urbana: University of Illinois Press. Weinacht, Stefan, and Ralf Spiller. 2014. Datenjournalismus in Deutschland. Publizistik 59: 411–33. [CrossRef] Wright, Scott, and Kim Doyle. 2019. The Evolution of Data Journalism: A Case Study of Australia. Journalism Studies 20: 1811–27. [CrossRef]

Wright, Kate, Rodrigo Zamith, and Saba Bebawi. 2019. Data journalism beyond majority world countries: challenges and opportunities. Digital Journalism 7: 1295–302. [CrossRef]

Young, Mary Lynn, and Alfred Hermida. 2015. From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. Digital Journalism 3: 381–97. [CrossRef]

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: 115–35. [CrossRef]

Zhang, Shuling, and Jieyun Feng. 2019. A Step Forward? Journalism Studies 20: 1281–300. [CrossRef]