

Trends in examining data journalism research: A systematic literature review in 2010-2023

Tren pemeriksaan terhadap penelitian-penelitian jurnalisme data: Tinjauan literatur sistematik tahun 2010-2023

M Fikri AR 

Department of Communication Science, Institute of Social Sciences, Sakarya University
Address: Sakarya Üniversitesi Esentepe Kampüsü, Serdivan, Sakarya, Turkey
E-mail: fikri.ar@ogr.sakarya.edu.tr

Article History: Received 21 November 2023; Accepted 22 July 2024; Published Online 29 July 2024

Abstract

Current studies of data journalism ideally produce knowledge about the use of data as news sources. The problem is that studies in data journalism are still partial; therefore, state of the art of data journalism is not well-known. This study aims to identify and analyze research topics, research methods, and frameworks in data journalism research in 2010-2023. Based on specific criteria, 76 primary studies were selected. This study uses a systematic literature review to analyze and answer the research questions that have been formulated. The results show that data journalism research focuses on topics: practical studies, theoretical studies, politics, education, visualization, storytelling, data journalism and new media, objectivity, sports, innovation, and crime reporting. The research methods are interviews, content analysis, case studies, mixed methods, surveys, ethnography, action research, thematic analysis, discourse analysis, grounded, and experiments. The frameworks identified are the Veglis and Maniou framework, Borges-Rey framework, and Lewis framework. This study concludes the topic of practical studies appears dominant because data journalism is a new phenomenon so its practices are important to research. The interview method is mostly used because it is familiar to researchers from various disciplines. The identification of these three frameworks shows the limited systematic study of data journalism.

Keywords: data journalism; framework identified; new phenomenon; research topic; systematic literature review

Abstrak

Studi terkini jurnalisme data idealnya memproduksi pengetahuan tentang penggunaan data sebagai sumber berita. Permasalahannya studi terkait jurnalisme data masih parsial sehingga gambaran state of the art jurnalisme data belum banyak diketahui. Studi ini bertujuan untuk mengidentifikasi dan menganalisis topik penelitian, metode penelitian, dan framework penelitian-penelitian jurnalisme data pada tahun 2010–2023. Berdasarkan kriteria yang spesifik didapatkan 76 primary studies yang dipilih untuk diteliti. Studi ini menggunakan metode systematic literature review untuk menganalisis bukti-bukti penelitian yang tersedia untuk menjawab pertanyaan penelitian yang telah dirumuskan. Hasil studi ini menunjukkan penelitian-penelitian jurnalisme data berfokus pada topik: kajian praktik, kajian teoritik, politik, pendidikan, visualisasi, storytelling, jurnalisme data dan media baru, objektivitas, olahraga, inovasi, dan pemberitaan kriminal. Adapun metode penelitiannya adalah wawancara, analisis isi, studi kasus, mix methods, survei, etnografi, action research, analisis tematik, analisis wacana, grounded, dan eksperimen. Framework yang teridentifikasi yaitu framework Veglis dan Maniou, framework Borges-Rey, dan framework Lewis. Studi ini menyimpulkan topik kajian praktik muncul dominan karena jurnalisme data merupakan fenomena baru sehingga dinamika praktiknya penting untuk diteliti. Adapun metode wawancara terbanyak digunakan karena familiar bagi peneliti dari berbagai disiplin ilmu. Teridentifikasinya tiga framework inilah yang menunjukkan masih terbatasnya studi sistematis terhadap jurnalisme data.

Kata Kunci: jurnalisme data; identifikasi framework; fenomena baru; topik penelitian; tinjauan literatur sistematik

Introduction

Data journalism is a genre of journalism where the news comes from the data (Chiumbu & Munoriyawa 2023). Hence, data journalism is also a process of extracting the essence of useful information of which the sources are data (Veglis & Maniou 2018). In academic history, after computers and the internet became more widespread in society, data journalism continued to develop in society. There are several terms that are used to describe data journalism, including: computer assisted reporting, precision journalism,

data driven journalism, computational journalism, programming journalism, and algorithmic journalism (Borges-Rey 2020). Data journalism as a further development of computer-based journalism practices has the advantage of making the news more open, transparent and accessible because the news content is based on data (Alieva 2023). However, because it is related to big data on the internet, not all journalists are able to develop data journalism at a practical level. At the practical level, data journalism requires quite specific skills related to investigative journalism methods, data management, and design for news visualization (Appelgren & Linden 2020).

From this reality, it can be seen that data journalism has expanded throughout the world and is described as the future of journalism (Wright & Nolan 2023). On a practical level, it is not simple. The stages in data journalism can be done by summarizing the data, organizing data in groups, counting and adding items, and checking some trends in the average (Richards 2023), then presenting it in the form of an attractive visualization. However, journalists usually experience difficulties in the process of making and visualizing the data in the form of journalistic work that is exciting and also easy to understand (Wright et al. 2019). These difficulties include practices related to the use of certain technology or software to process and visualize data in an attractive form. This difficulty means that there is still a dichotomy between the fields of journalism, data science and the computer field. In practice, data journalism is also influenced by various contextual factors such as politics, culture, money and power (Gray & Bounegru 2021).

Along with the development of communication and information technology, data journalism has now been practiced in various countries, such as in Africa (Chiumbu & Munoriyarwa 2023), Spanish and Austrian regions (Cordoba-Gabus et al. 2023), China (Wu 2022), Sweden (Widholm & Appelgren 2022), the United States (Clark & Rodriguez 2021), Australia (de-Lima-Santos et al. 2021), Arab countries (Fahmy & Attia 2021), the United Kingdom (Borges-Rey 2020), and Pakistan (Jamil 2019). However, even though it has been implemented in various countries, it turns out that the field of data journalism still lacks of theoretical studies. Especially in a literature review that summarizes the theoretical and the practical studies of data journalism.

Therefore, the basic understanding of how to obtain data, understanding data, and delivering data still needs to be researched (Gray et al. 2012). More specifically, the skills which are related to the ability to processing the data, programming data, visualizing data, cleaning data, develop data-driven journalism applications, check data reliability, develop news from data arrangement, obtaining data, including integrating and publishing in the content management system (Heravi 2019). Some of these skills make data journalism quite complex for journalists, so that not all media companies are able to implement data journalism optimally. However, the trend in the application of data journalism appears in line with the increasingly intensive use of data by various companies or institutions in the fields of economics, education, politics and health.

Compared to other types of journalism, the advantage of data journalism is more attractive and easier to understand. Data journalism helps people to understand quantitative information more easily (Alieva 2023). More than that, the use of data as a news source makes data journalism more objective and transparent (Wu 2022), compared to the application of conventional journalism where the basis of the news is statements or opinions from the sources. Recent studies show that the development of media coverage throughout the world shows the high demand of society for the presence of news that is carried out using data that are displayed in an attractive form and easy to understand (Wang et al. 2023). Hence, more studies on data journalism that look at the potential and challenges associated with its application need to be developed.

Recent studies related to data journalism can help journalists, students and researchers who see that data journalism has the potential to be applied as a journalism model that is more transparent, objective (Pentzold & Fechner 2021), improves the knowledge of modern journalism (Neyazi & Schroeder 2021) and also the quality of democracy in a country (Parasie 2022). Therefore, the research on data journalism, especially those related to its epistemological aspects, is an important area in the field of journalism (Schaetz et al. 2023). In practice, quite a lot of research related to data journalism has been published, but the research topics tend to be separate and partial. For this reason, a comprehensive mapping of current studies on data journalism is needed.

The contribution of knowledge in this study is a scientific and systematic explanation of research trends in the field of data journalism related to research topics, research methods, and frameworks. This study also shows that there are challenges faced by data journalism that are different from conventional journalism in general, because data journalism requires skills in the fields of journalism and computer science. In other words, journalists in the present and future need to equip themselves with media technology skills. The novelty of this study is in the results which present the latest topics, research methods and frameworks that are growing up in the field of data journalism.

Research Method

This study uses a systematic literature review method to comprehensively understand research topics, research methods and frameworks in data journalism research in 2010-2023. The discussion in this section includes review method, research questions, search strategy, study selection, data extraction, as well as study quality assessment and data synthesis, which are attempted to be as systematic as possible to answer the formulated research problems. A more complete description about it is presented in Figure 1.

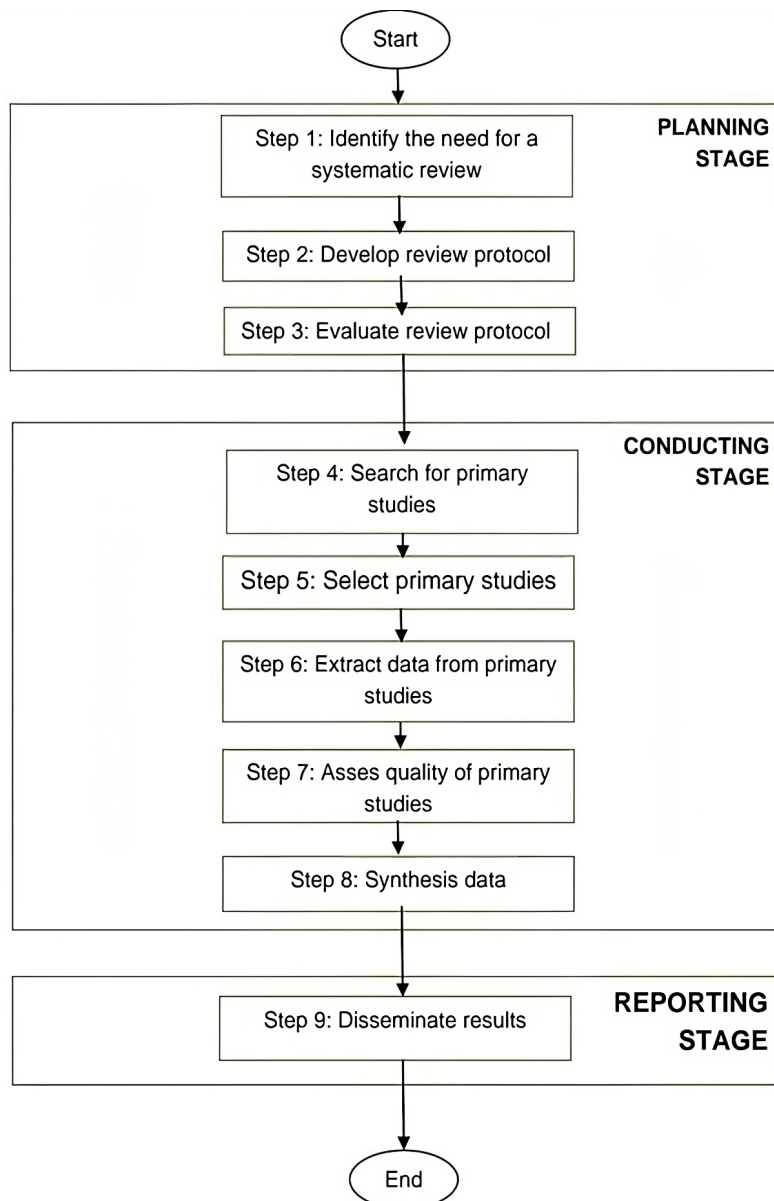


Figure 1.
Systematic literature review steps
Source: Compiled by the author

In the field of social sciences, systematic literature review method has developed rapidly. Systematic reviews here attempt to comprehensively identify, assess, and synthesize all relevant studies on a particular topic (Petticrew & Roberts 2006). Systematic literature review has several advantages over traditional reviews such as the presence of several specific procedures. The process is carried out using extensive search methods, previously conducted search strings and inclusion and exclusion criteria that enable researchers to obtain relevant data (Shaffril et al. 2020). In this case, a systematic literature review is carried out to study the primary studies related with data journalism research.

As shown in Figure 1, the systematic literature review was carried out in three stages: planning stage, conducting stage, and reporting stage. In the first step, the requirements for a systematic literature review are identified (Step 1). Then, systematic literature review materials related to data journalism are identified and collected. The review protocol is designed to carry out and execute the review process and reduce the possibility of bias perceptions. At this stage, the research question is defined, then the research strategy, the process of selecting primary studies with exclusion and inclusion criteria, quality of checking, and finally the process of data synthesis and presenting conclusions are performed. In its implementation, the review protocol continues to be developed and evaluated, repeatedly until mature. A systematic description of the steps in this systematic literature review is shown in Figure 1.

The research questions (RQ) were specified to maintain the focus of this systematic literature review. This part is designed by explaining Population, Intervention, Comparison, Outcomes and Context (PICOC) or certain criteria in the literature review (Petticrew & Roberts 2006). Table 1 shows the PICOC structure of the research questions that have been prepared.

Table 1.
Summary of PICOC

Population	Data journalism, digital media, data-driven journalism, computational journalism.
Intervention	Getting data, understanding data, delivering data, data visualization, transparency, objectivity, public sphere, politics.
Comparison	Data journalism and traditional journalism.
Outcomes	Effective data management, more objective news, news based on data, news visualization in interesting form.
Context	Media studies in academia and industry, news media that implement data journalism.

Source: Compiled by the author

Furthermore, the research questions and motivation for conducting the analysis in this systematic literature review are shown in Table 2. From the primary studies that have been collected, data regarding research topics, research methods and frameworks used in data journalism are then extracted to answer RQ1 to RQ6. More specifically, RQ1 is a question formulation to find out the most significant journals in the data journalism study, RQ2 is to find out the most active researchers in the data journalism study, RQ3 is to find out research topics that are emerge in the data journalism study, RQ4 and RQ5 are to find out the methods that can be and are most used in the data journalism study, and RQ6 seeks to find out a potential framework for developing the data journalism studies. Figure 2 shows the basic mind map of this systematic literature review. The main aim of this systematic literature review is to identify research topics, research methods, and frameworks that have potential to develop data journalism as a field study.

Table 2.
Research questions on literature review

ID	Research Question	Motivation
RQ 1	Which journal is the most significant data journalism journal?	Identify the most significant journals in the data journalism field.
RQ 2	Who are the most active and influential researchers in the data journalism field?	Identify the most active and influential researchers in the data journalism field.
RQ 3	What kind of research topics are selected by researchers in the data journalism field?	Identify the research topic and trends in the data journalism field.
RQ 4	What kind of methods are used for data journalism?	Identify the methods that are used in the data journalism field.
RQ 5	What kind of methods are used most often for data journalism?	Identify the most used methods for data journalism field.
RQ 6	What kind of frameworks are proposed for data journalism?	Identify the frameworks that have potential to develop the study of data journalism.

Source: Compiled by the author

The search process (step 4) in this systematic literature review consists of several activities, such as selecting a digital library, defining the search string, executing a pilot search, refining the search string and storing a list of names of primary studies in the digital library that match to the search string. Before starting the search, a series of appropriate databases were selected to be able to retrieve relevant articles.

The most popular literature databases in the study area were searched to obtain a series of relevant studies. A comprehensive perspective is needed to obtain appropriate literature. Several lists of digital databases in this systematic literature review are as follows: IEEE Explore (ieeexplore.ieee.org), ScienceDirect (sciencedirect.com), Springer (springerlink.com), Scopus (scopus.com), and Google Scholar (scholar.google.com)

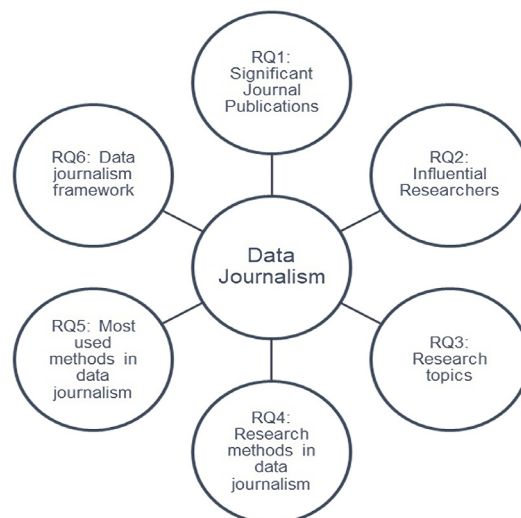


Figure 2.
Basic mind map of the systematic literature review on data journalism
Source: Compiled by the author

To find primary studies that are relevant to the purpose of writing this systematic literature review, the search string needs to be developed according to the following steps: (1) Identify search terms from PICOC, especially from population and intervention; (2) Identify search terms from research questions; (3) Identify the search string from relevant titles, abstracts, and keywords; (4) Identification of synonyms, alternative pronunciations, and antonyms of search terms; (5) Build a sophisticated search string using search themes, AND, or OR. The search string used in this systematic literature review is: (data journalism OR computational journalism OR data driven journalism) AND (digital media OR news media OR online media) AND (public sphere OR digital public sphere OR political public sphere) AND (objectivity OR factuality OR Impartiality).

Settings to the search string are made, but the original is kept. Setting the search to the search string will dramatically increase the extensive list of irrelevant studies. The search string is set according to the specific statements of each database. The database is searched by title, keywords and abstract. The search was limited to publications in the years 2010-2023. The type of publication sought was journal paper, while conference proceedings were not included. Searches are limited to English languages only. Inclusion and exclusion criteria were used to select primary studies. A more complete explanation about these criteria is shown in Table 3.

Details of the search process and numerical results of the studies identified from each step are shown in Figure 3. As shown in Figure 3, the selection process (step 5) is carried out in two steps, namely: (1) The exclusion step from primary studies is carried out based on the title and abstract and (2) The inclusion step from primary studies is based on the entire text. Literature review studies and other studies that were not research results were excluded. Primary studies which are the results of research in the area of data journalism studies are included in this systematic literature review.

The final list of primary studies selected for the first stage consisted of 1499 primary studies. Then, the text of the primary studies was analyzed based on the title and abstract until 217 primary studies were obtained. Next, referring to the exclusion and inclusion criteria, analysis was carried out on primary studies to obtain 105 primary studies. In the next stage, similar studies by the same author were removed. So finally, 76 primary studies were found remaining after the inclusion and exclusion criteria were carried out based on selection of all texts.

Table 3.
Inclusion and exclusion criteria

Inclusion Criteria	<p>Studies in academic and industrial fields that analyze the problem in data journalism field.</p> <p>Studies discussing and comparing the application of reporting in data journalism field.</p> <p>For studies that have conference and journal versions. Only the journal version will be included.</p> <p>For duplicate publications of the same study, only the most complete newest one will be included.</p> <p>Studies without strong validation in data journalism.</p>
Exclusion Criteria	<p>Studies that discuss topics, methods, and frameworks outside the topic of data journalism.</p> <p>Studies not written in English.</p>

Source: Compiled by the author

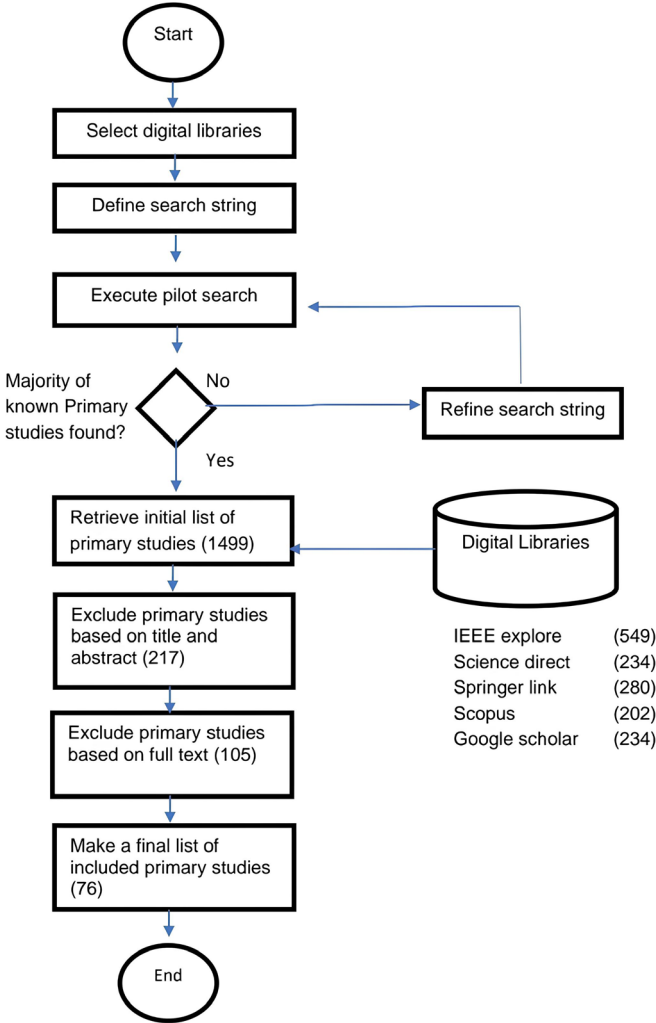


Figure 3.
 Search and selection of primary studies
 Source: Compiled by the author

Selected primary studies were extracted to obtain data that could contribute to answering the research questions in this systematic literature review. For each of the 76 selected primary studies, properties for data extraction were completed (step 6). In general, the data extraction steps are designed to collect data from existing primary studies to answer research questions. Properties are identified through research questions, after which analysis is carried out. The properties used to answer the research questions are shown in Table 4. Data extraction is carried out iteratively or continuously.

Table 4.
 Data extraction properties mapped to research question

Property	Research Question
Research and Researcher	RQ1, RQ2
Research topics	RQ3
Research method	RQ4
Most used method in data journalism	RQ5
Framework in data journalism	RQ6

Source: Compiled by the author

The study of quality assessment (step 7) was used to guide the process of interpreting the synthesis findings and to define the strength of the analysis and draw conclusions. The purpose of data analysis is to aggregate the findings of the selected studies and to answer the research questions. One piece of evidence found may have weaker strength, but the aggregation of a lot of data can strengthen the points of the findings. Data extraction in this systematic literature review includes quantitative data and qualitative data. Next, the data are tabulated in a way that is suitable with the research question, which is then developed by the researcher into a data synthesis to answer the research question. Several visualization tools such as tables, charts, and pie charts are used to present the distribution of research methods used in the data journalism study.

Results and Discussion

After explained in the initial part the importance of researching the field of data journalism as a new field of journalism that is developing with the presence of big data on the internet, then also explaining the method aspects in this study, it becomes clearer that the field of data journalism brings hope regarding the presence of a more objective type of journalism and has great potential for further development. An explanation of this will be presented systematically in the results and discussion section which consists of significant journal publications, most active and influential researchers, research topics in data journalism field, methods in data journalism field, and frameworks in data journalism field, then continued with discussion of the research findings.

Significant journal publications

In this systematic literature review, a total of 76 primary studies were analyzed. The distribution of publication years is presented to show how researchers' attention to the field of data journalism has developed over the years. An overview of the distribution of data journalism studies over several years is shown in Figure 4. The existence of studies published since 2010 shows that the study of data journalism is a current topic that is receiving increasing attention. In this case, it should be noted that the study of data journalism has become more advanced since 2014, especially after the publication of *The Data Journalism Handbook* in 2012 which encouraged researchers to examine the field of data journalism more empirically. It can be understood that Figure 4, apart from illustrating the development of studies on data journalism, also shows that the field of data journalism remains relevant to study today.

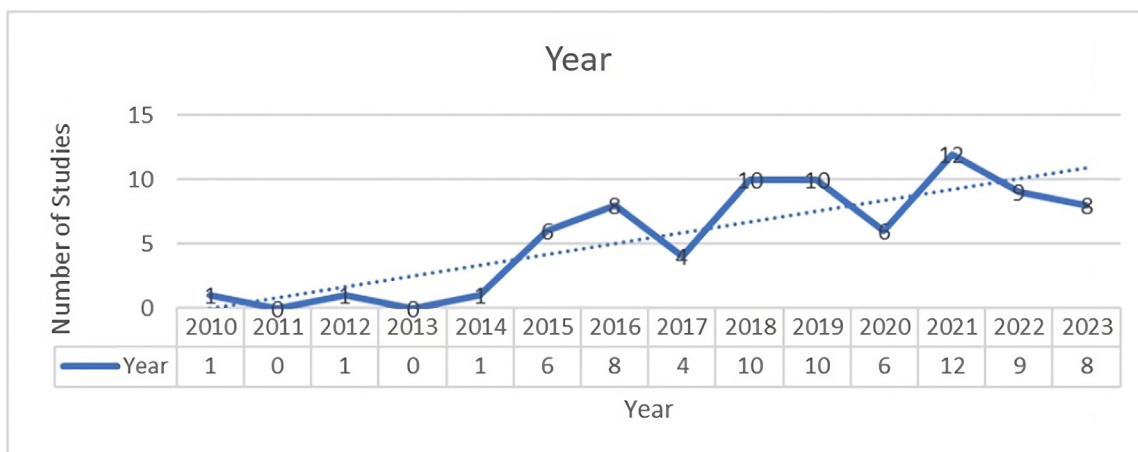


Figure 4.
Distribution of selected studies over the years
Source: Compiled by the author

Based on the selected primary studies, the most important journals in the field of data journalism are shown in Figure 5. Please note that conference proceedings are not included in this figure.

AR: "Trends in examining data journalism research"

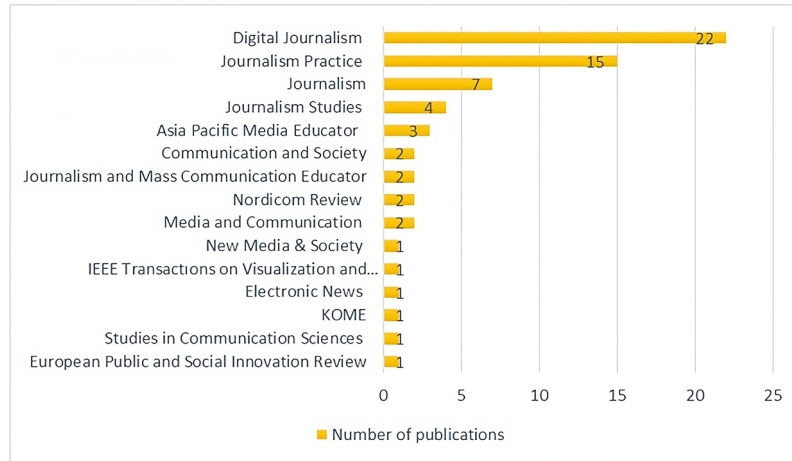


Figure 5.
Journal publications and distributions of selected studies
Source: Compiled by the author

Table 5.
Scimago journal rank (SJR) of selected journals

No	Journal Publications	SJR	Q Category
1	Digital Journalism	2.63	Q1 in Communication
2	New Media & Society	2.08	Q1 in Communication
3	Journalism Studies	1.59	Q1 in Communication
4	Journalism	1.51	Q1 in Communication
5	Journalism Practice	0.99	Q1 in Communication
6	Media and Communication	0.87	Q1 in Communication
7	Communication and Society	0.43	Q1 in Cultural Studies
8	Nordicom Review	0.5	Q2 in Communication
9	Journalism and Mass Communication Educator	0.39	Q2 in Communication
10	KOME	0.3	Q2 in Communication
11	Journal of Content, Community & Communication	0.25	Q3 in Communication
12	Asia Pacific Media Educator	0.24	Q3 in Communication
13	Studies in Communication Sciences	0.23	Q3 in Communication
14	Central European Journal of Communication	0.19	Q3 in Communication
15	European Public and Social Innovation Review	0.14	Q4 in Sociology and Political Science

Source: Compiled by the author

Then, based on the selected primary studies, Table 5 shows the Scimago Journal Rank (SJR) value and Q categories (Q1-Q4) of the most important journals in the field of data journalism. Journal publications are arranged in a table based on their SJR values.

Most active and influential researchers

From the primary studies that have been selected, researchers who have contributed and are active in the field of data journalism research have been identified. Table 6 shows the most active and influential researchers in the data journalism field. Researchers are listed based on the number of studies in this literature review.

Table 6.
Influential researchers and number of studies

No	Researchers	Number of Studies
1	Mathias-Felipe de-Lima-Santos	5
2	Ester Appelgren	5
3	Mary Lynn Young	4
4	Shangyuan Wu	4
5	Norman P Lewis	3
6	Florian Stalph	3
7	Alfred Hermida	3
8	Eddy Borges-Rey	3
9	Bahareh Heravi	3
10	Sylvain Parasio	2
11	Wibke Weber	2
12	Ramon Salaverria	2
13	Sergio Splendore	2
14	Geeta Kashyap	2
15	Harikrishnan Bashkaran	2

Source: Compiled by the author

Research topics in data journalism field

Basically, data journalism is a significant research area in the field of journalism and communication science or social sciences. Therefore, knowing the development of research topics in data journalism field also useful for developing communication science or social sciences in general. Analysis of selected primary studies reveals that the latest research on data journalism focuses on several topics.

First, practical study. This topic refers to practical or applied issues in data journalism. In this systematic literature review, there are 21 primary studies that examine the topic of practical studies in data journalism field. For example, research conducted by Jamil (2019), Vrabec & Proner (2021), Haim (2022), and de-Lima-Santos & Mesquita (2023).

AR: "Trends in examining data journalism research"

Second, theoretical study. The purpose of this topic examines the theoretical aspects of data journalism which continue to develop over time. In this systematic literature review, there are 18 primary studies that examine theoretical study topics such as conducted by Baack (2018), de-Lima-Santos & Salaverria (2021), Hannaford (2022), and Morini (2023).

Third, politics. The topics about politics are topics in data journalism that are related to political events such as presidential election activities that occur in various regions. In this review there are nine primary studies that examine political topics such as research conducted by Lewis & Waters (2018), Alieva (2023), Chiumbu & Munoriyawa (2023) and Wu (2023).

Fourth, education. Educational topics here mean research topics in data journalism that are related to the data journalism education process held by various universities in various countries. In this review there are eight primary studies that examine this topic such as Gray, Gerlitz, & Bounegru (2018), Heravi (2019), Kashyap & Bashkaran (2020), and Lewis (2021).

Fifth, visualization. This topic focuses more on aspects of media display and news visualization which can be displayed in various different and attractive appearances for news readers. In this review, there are seven primary studies that examine this topic such as research conducted by Stalph and Heravi (2021) and Guan & Wang (2022).

Sixth, storytelling. This topic examines the storytelling techniques used by the media to deliver data to audience. This topic is close to the topic of visualization, but the topic of storytelling is broader since it relates to aspects of obtaining data, processing data, and delivering data. In this review, there are three primary studies that examine this topic, such as research conducted by Ojo & Heravi (2018), Pentzold & Fechner (2021), and Auvaart (2022).

Seventh, data journalism and new media. This topic is related to the phenomenon of data journalism which is related to the use of new media to support the data journalism practice. In this review, there are two primary studies that examine the topics of data journalism and new media, namely Appelgren (2016) and Li (2022).

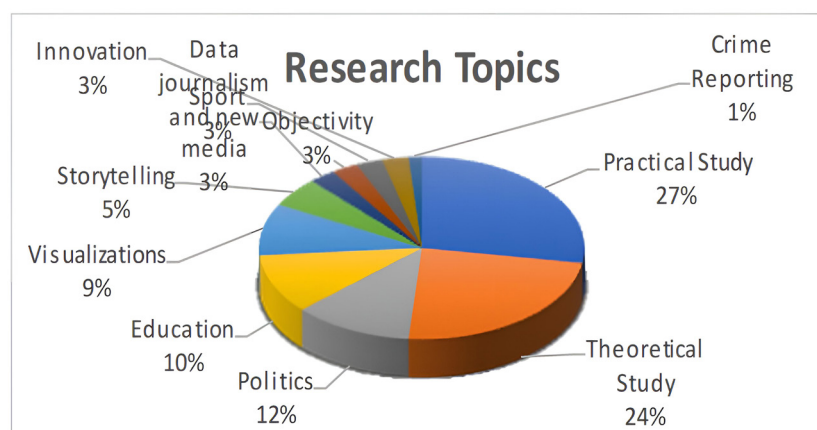


Figure 6.
Distribution of research topics
Source: Compiled by the author

Eighth, objectivity. This topic discusses the issue of objectivity in data journalism as a type of journalism that developed in society. In this context, the objectivity is one of the advantages of data journalism because the news in it comes from data, not from statements or opinion of news sources. In this systematic literature review, there are two primary studies that examine the topic of objectivity, namely Wu (2021) and Wu (2022).

Ninth, sports. This topic is related to sports phenomena which are the subject of media coverage. Generally, sport news continues to get attention from media companies to spread sporting events because of the appeal of sports as entertainment, and often sports are also an opportunity to unite people's sense of nationalism in society. In this study, there are two primary studies that examine sports topics, namely Horky & Pelka (2017), and Vural & Masip (2021).

Tenth, innovation. This topic is related to the new efforts and new practices to develop the data journalism carried out by academics and practitioners. There are two primary studies that examine this topic, namely Zhang & Chen (2020) and de-Lima-Santos & Mesquita (2021).

Eleventh, crime reporting. This topic is related to criminal acts that are highlighted as news. In this review, there is one primary study which examines the topic of crime reporting, namely Young & Hermida (2015). For more detailed information about the research topics in this systematic literature review, see the distribution of research topics presented in Figure 6.

Methods in data journalism field

From research on data journalism conducted since 2010, 11 methods were identified that were used for research, namely: Interview, content analysis, case study, mixed methods, survey, ethnography, action research, thematic analysis, discourse analysis, grounded, and experiment. A more complete picture of the 11 methods used in the field of data journalism is shown in Figure 7 and Table 7.

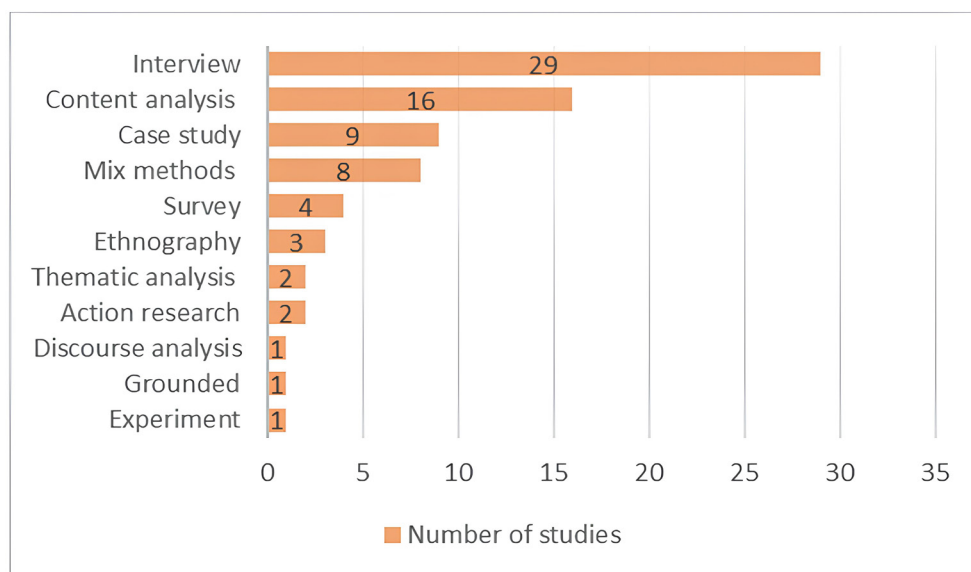


Figure 7.
Method used in data journalism field
Source: Compiled by the author

From the 11 methods used by researchers above, the six methods most widely used in the field of data journalism have also been identified. These methods are: interview (29), content analysis (16) case study (9), mixed methods (8), survey (4), and ethnography (3). Meanwhile, other methods such as discourse analysis, grounded and experiment are still not widely used, so further research is still needed to develop studies in data journalism field. For a more complete explanation of the methods used in data journalism, see the Table 7.

Table 7.
Distribution of methods in the research of data journalism

No	Methods	Amount	Percentage
1	Interview	29	38,15
2	Content analysis	16	21,05
3	Case study	9	11,84
4	Mixed methods	8	10,52
5	Survey	4	5,26
6	Ethnography	3	3,94
7	Action research	2	2,63
8	Thematic analysis	2	2,63
9	Discourse analysis	1	1,31
10	Grounded	1	1,31
11	Experiment	1	1,31

Source: Compiled by author

Frameworks in data journalism

The framework in this systematic literature review places more emphasis on research on data journalism applied in the news media. From the review process of the 76 existing primary studies, several basic things can be understood. First, there are three elements that form the data journalism, namely: (a) data as a news source; (b) data journalists as actors in the media and data journalism as a reporting technique; and (c) the public as recipients of the news. Second, the general process of implementing data journalism also includes three stages, namely: (a) obtaining data, (b) understanding the data, and (c) delivering the data to the public. Third, data journalism as a theory of knowledge and tools is actually neutral, depending on the actor who places data journalism in a particular context, so that data journalism has specific characteristics or is different from one another, but actually has basic similarities in the process.

In this systematic literature review, three frameworks were identified that describe data journalism in their specific contexts, namely: the Veglis & Maniou Framework (2018), the Borges-Rey Framework (2020), and the Lewis Framework (2021). A more complete explanation is provided below:

The Veglis and Maniou framework

This framework was developed in the context of applying the two-step flow of communication theory. This framework describes the position of data journalism as the party that translates complex data to the public. The position of data journalism is in the middle between the data source and the recipient or public. As shown in Figure 8, the main elements of this framework are: (1) data as a source, (2) journalist data as an intermediary whose position is in the middle, and (3) the public as a recipient.

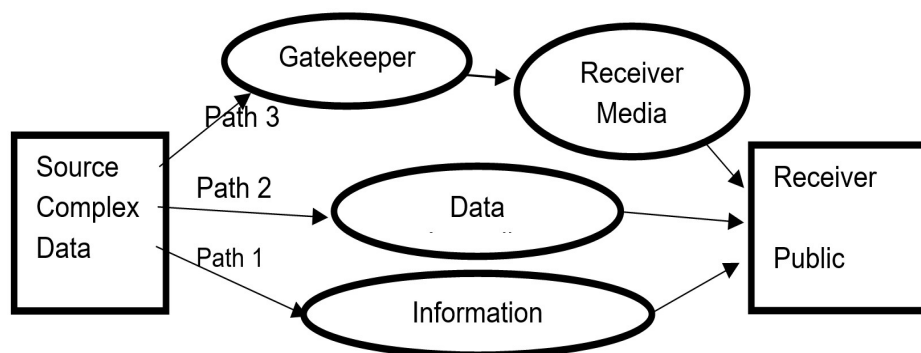


Figure 8.
The mediated data model of communication
Source: Adapted from Veglis & Maniou (2018)

According to this framework, there are three process paths for messages from data sources to the public as recipients of information. Path 1 shows the existence of an “information seeker” who acts as an information bridge between data sources and the public. In path 1, information seekers do not easily get the essence of the message from the data source even though they can find it themselves. Path 2 shows the existence of data journalists as an information bridge between complex data sources and the public. Path 2 also shows that data journalists play the role of extracting the essence of information from data sources using data journalism techniques and disseminating it to the audience with attractive visualizations. Path 3 shows the existence of a “gatekeeper” or “communication specialist” who filters and processes information which is then conveyed through new media and then distributed to the public.

According to Veglis and Maniou most information flows through path 3. Path 2 is used for several cases of searching for certain information, some of which have appeared in path 3. Meanwhile, route 1 is rarely used and is mostly carried out by media stakeholders who are not traditional media users such as bloggers, alternative media activists, and citizen journalists, and has a smaller impact on society.

The Borges-Rey framework

The Borges-Rey (2020) framework was developed using the conceptual lens of Materiality, Performativity & Reflexivity. This framework developed the thesis that data journalism developed in Scotland, Wales and Northern Ireland (UK) has unique practices that deserve to be given its own epistemological framework. The Borges-Rey framework was developed based on several studies on the epistemology of data journalism that had been conducted previously by Coddington et al. (Borges-Rey 2020).

This framework describes data journalism as being in the pull between practices with a traditional newsroom approach or newshound and a computational logic approach or techie. The newshound approach is understood as a traditional journalistic way of handling and interacting with data or in a simpler sense that can be understood as a traditional newsroom approach. Meanwhile the techie approach refers to an emerging journalistic approach to data, based more on computational logic and mindset. This approach is more advanced because it is closer to the academic approach and entrepreneurial approach.

In the context of data journalism in Scotland, Wales and Northern Ireland, according to Borges-Rey, materiality is closer to the newshound approach. Then, in terms of performativity, it tends to be techie because of the scarcity of resources and also relies on collaboration and professionalism. Meanwhile, reflexivity tends to the techie approach because the stage of journalism continues to develop and there is an appreciation for technology in journalism. A more complete description of the Borges-Rey framework can be seen in Figure 9.

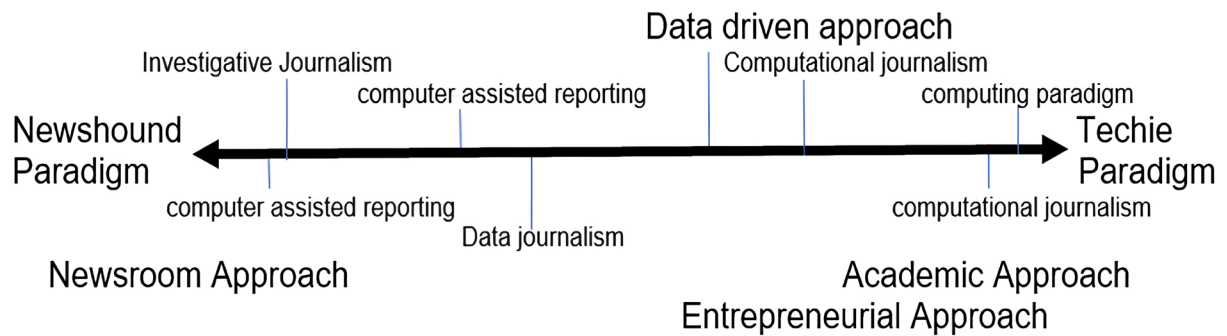


Figure 9.
The framework of epistemology in data journalism practice
Source: Adapted from Borges-Rey (2020)

The Lewis framework

The Lewis (2021) framework describes data journalism, especially in the field of education. This framework was developed from a thematic evaluation of data journalism courses at universities, resulting in a typology that outlines the data journalism process. As shown in Figure 10, this framework explains the data journalism process, starting with data acquisition and cleaning the data, then there is a pattern detection process, then the data representation process. This framework describes data journalism in the field of education. It can be understood starting from building data literacy capacity, then (1) data acquisition and cleaning, (2) pattern detection, and (3) data representation.

The framework developed by Lewis (2021) increases academic understanding that data journalism is different from peripheral technology, and also helps identify coding skills as a supporting skill. This also encourages the emergence of a new definition of data journalism that data can be used as a journalistic tool to detect patterns, or for visual representations based on numerical evidence, thereby allowing audiences to distinguish patterns and understand the reality of events better. A more complete description of the Lewis (2021) framework can be seen in the Figure 10.

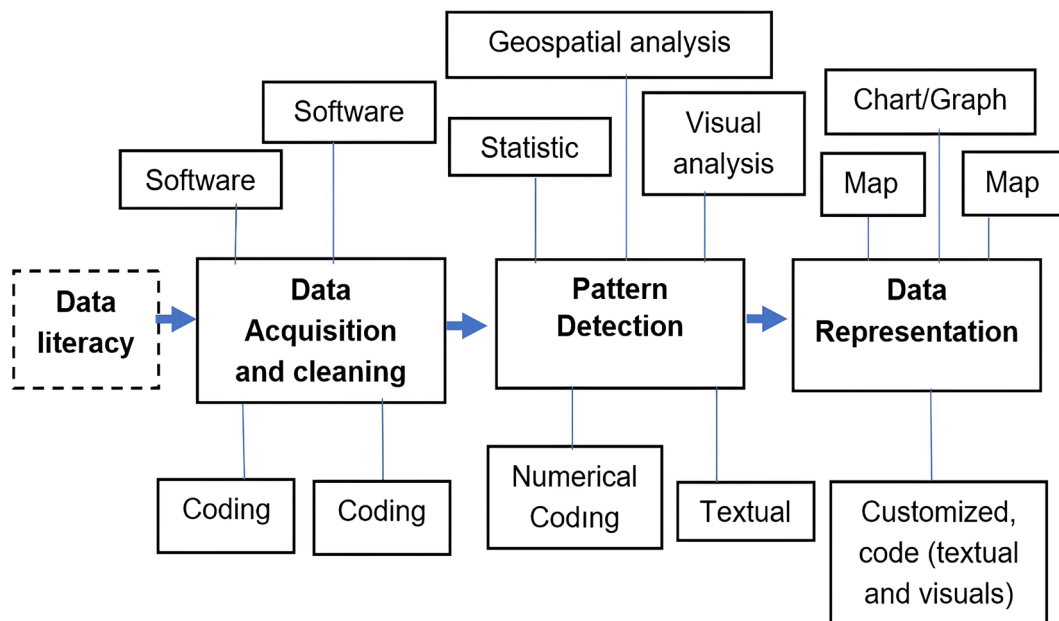


Figure 10.
Framework for data journalism as a field for education
Source: Adapted from Lewis (2021)

The research results presented above show that the framework being developed in the field of data journalism is still very limited. This is because data journalism is a relatively new field of study and has not received much attention, especially in the field of theoretical studies. In this context, Veglis and Maniou's framework tends to show data journalism as an intermediary or translator between big data and society. Then, the Borges-Rey framework shows data journalism in the context of the dynamics of journalism in England (UK), while the Lewis framework shows data journalism in the field of education. Meanwhile, the data journalism framework in the field of political reporting, conflict phenomena or culture has not yet emerged, so further research is still needed to complement existing frameworks.

In the context of the use of research methods, the interview method is most widely used as a method for obtaining data in data journalism, indicating that this method is familiar and affordable for data journalism researchers, while discourse analysis, grounded, and experimental methods are the least used because they tend to be more complex and it takes longer time to research some topics in the field of data journalism. On the other hand, the challenge of the dichotomy between communication science or social science with computer science is also a challenge in the area of methodology for researchers who are interested in exploring the field of data journalism, because a comprehensive understanding of data journalism requires good mastery in journalism and also in computer science.

Furthermore, the topic of practical studies is the most researched topic in the field of data journalism because data journalism is still considered a new phenomenon, so the practice and application of data journalism is considered important and interesting to research. Apart from that, data journalism is also considered to bring hope for the emergence of a more objective model of journalism, because the source of the news is data, not opinions or statements from certain figures.

Then, the topic of crime is still few to be researched because there is little data about crime available in society. To create crime news based on data, journalists need to compile separate crime data from various sources, and this requires a lot of time and energy. Apart from that, compiling these data also requires competence in the field of data processing, which is not simple, so that crime news does not appear too much, which results in a lack of research topics on crime.

In addition, the fact that practical study topics dominate research topics in the field of data journalism also indicates that the practice of data journalism is growing quite rapidly in the media industry. This shows that news based on data is quite popular and awaited by the public. There is hope that objective news based on data can grow and becomes an alternative to news that originates from statements by certain figures. At the same time, the various methods that can be used to research data journalism are also good news for researchers in that the field of data journalism is a new area that is increasingly open to continuous research. Moreover, the limited frameworks developed in data journalism show that there is still a need for various frameworks in the field of data journalism.

This systematic literature review contributes to knowledge by showing various research topics, research methods and frameworks that are visible for use and development in the field of data journalism. Meanwhile, existing studies pay less attention to comprehensive or holistic analysis related to strengthening theoretical studies of data journalism, regarding research topics, research methods, and research frameworks that have the potential for further research. Previous studies tend to examine data journalism in its specific context, such as studies on the dichotomy of commercial media and public service media organizations (de-Lima-Santos et al. 2021), investigative studies of data journalism and its collaboration opportunities (Uretta & Rodriguez 2021), and study of the use of softnews and hardnews in data journalism (Widholm & Appelgren 2022). This study also shows that there are challenges faced by data journalism that are different from conventional journalism in general, because data journalism requires skills in the fields of journalism and computer science. This means that journalists in the present and future need to equip themselves with media technology skills in their daily coverage.

Conclusion

This systematic literature review has made several findings. First, analysis of the selected primary studies states the latest data journalism research focuses on practical studies, theoretical studies, politics, education, visualization, storytelling, data journalism and new media, objectivity, sports, innovation, and crime reporting. Second, the research methods used in data journalism research are interviews, content analysis, case studies, mixed methods, surveys, ethnography, action research, thematic analysis, discourse analysis, grounded, and experiments. Third, the finding of this study also identified three frameworks describing the dynamics of data journalism in various contexts, namely the Veglis and Maniou framework, the Borges-Rey framework, and the Lewis framework. These frameworks have specific context because the Veglis and Maniou framework describes data journalism in the context of two-step flow of communication theory, the Borges-Rey framework describes data journalism in the context of journalism epistemology in England (UK), and the Lewis framework describes data journalism in the context of education.

The systematic literature review method developed in this study allows for the discovery of space for systematic explanations about research topics, research methods and research frameworks that have the potential to be developed in data journalism studies. With this, it can also be understood that the development of data journalism occurred because of the abundant presence of big data on the internet. This has the consequence that journalists' abilities need to be improved toward the ability to read, understand and process data on the internet. This also suggests the importance of increasing academic competence in the fields of journalism and computer science for students, researchers or journalists, who want to explore the field of data journalism.

This study has limitation in that it examines primary studies related with data journalism from 2010 to 2023. Therefore, the perspective contained in this study is still limited to an abstract theoretical viewpoint and an introduction perspective for further research. This means that further research is still needed that examines the dynamics of data journalism in various contexts. For example, from the perspective of journalist perceptions of the potential for implementing data journalism or the perspective of media companies that are trying to implement data journalism as a reporting standard that journalists need to develop. In other words, this research recommends further qualitative or quantitative research that examines the dynamics of the development of data journalism in a more empirical context. It is hoped that the research results will complement the research findings carried out by this study.

References

- Alieva L (2023) How American media framed 2016 Presidential election using data visualization: The case study of the New York Times and the Washington Post. *Journalism Practice* 17 (4):814-840. <https://doi.org/10.1080/17512786.2021.1930573>.
- Appelgren E & Linden CG (2020) Data journalism as a service: Digital native journalism expertise and product development. *Media and Communication* 8 (2):62-72. <https://doi.org/10.17645/mac.v8i2.2757>.
- Appelgren E (2016) Data journalist using facebook: A study of a resource group created by journalist, for journalist. *Nordicom Review* 37 (1):1-14. <https://doi.org/10.1515/nor-2016-0007>.
- Auvaart L (2022) Fighting COVID-19 with data: An analysis of data journalism projects submitted to Sigma Awards 2021. *Central European Journal of Communication* 3 (32):379-395. [https://doi.org/10.51480/1899-5101.15.3\(32\).3](https://doi.org/10.51480/1899-5101.15.3(32).3).
- Baack S (2018) Practically engaged: The entanglements between data journalism and civic tech. *Digital Journalism* 6 (6):673-692. <https://doi.org/10.1080/21670811.2017.1375382>.
- Borges-Rey E (2020) Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism* 21(7):915-932. <https://doi.org/10.1177/1464884917693864>.

- Chiumbu S & Munoriyarwa A (2023) Exploring data journalism practices in Africa: Data politics, media ecosystems and newsroom infrastructures. *Media, Cultures & Society* 1-18. <https://doi.org/10.1177/01634437231155341>.
- Clark MA & Rodriguez J (2021) Big data and journalism: How American journalism is adopting the use of big data. *Novum Jus* 15 (1):69-89. <https://doi.org/10.14718/NovumJus.2021.15.1.4>.
- Cordoba-Gabus A, Huber A, & Farias-Batlle P (2023) Data journalism in Spain and Austria: Features, organizational structure, limitations, and future perspective. *Profesional de la Información* 32 (1):1-16. <https://doi.org/10.3145/epi.2023.ene.16>.
- de-Lima-Santos MF & Mesquita L (2023) Data journalism in Favela: Made by, for, and about forgotten and marginalized communities. *Journalism Practice* 17 (1):108-126. <https://doi.org/10.1080/17512786.2021.1922301>.
- de-Lima-Santos MF & Mesquita L (2021) Data journalism beyond technological determinism. *Journalism Studies* 22 (11). <https://doi.org/10.1080/1461670X.2021.1944279>.
- de-Lima-Santos MF & Salaverria R (2021) From data journalism to artificial intelligence: Challenges faced by La Nacion in implementing computer vision in news reporting. *Palabra Clave* 24 (3):1-40. <https://doi.org/10.5294/pacla.2021.24.3.7>.
- de-Lima-Santos MF, Schapals AK, & Bruns A (2021) Out-of-the-Box versus in-house tools: How are they affecting data journalism in Australia? *Media International Australia* 18 (1):152-166. <https://doi.org/10.1177/1329878X20961569>.
- Fahmy N & Attia MAA (2021) A field study of Arab data journalism practices in the digital era. *Journalism Practice* 15 (2):170-191. <https://doi.org/10.1080/17512786.2019.1709532>.
- Gray J, Gerlitz C, & Bounegru L (2018) Data infrastructure literacy. *Big Data & Society* 1-13. <https://doi.org/10.1177/2053951718786316>.
- Gray J & Bounegru L (2021) *The data journalism handbook: Towards a critical data practice*. Amsterdam: Amsterdam University Press.
- Gray J, Bounegru L, & Chambers L (2012) *The data journalism handbook: How journalist can use data to improve the news*. Sebastopol: O'Reilly.
- Guan T & Wang Q (2022) Examining the democratic potential of data journalism in curing misinformation. *Journalism Practice*. <https://doi.org/10.1080/17512786.2022.2124433>.
- Haim M (2022) The German data journalist in 2021. *Journalism Practice* 1-20. <https://doi.org/10.1080/17512786.2022.2098523>.
- Hannaford L (2022) The discourse of data journalism. *Journalism* 24 (11):1-21, <https://doi.org/10.1177/14648849221107803>.
- Heravi BR (2019) 3Ws of data journalism education: What, where and who? *Journalism Practice* 13 (3):349-366. <https://doi.org/10.1080/17512786.2018.1463167>.
- Horky T & Pelka P (2017) Data visualization in sport journalism: Opportunities and challenges of data driven journalism in german football. *Digital Journalism* 5 (5):587-606. <https://doi.org/10.1080/21670811.2016.1254053>.
- Jamil S (2019) Increasing accountability using data journalism: Challenges for the Pakistani Journalists. *Journalism Practice*: 1-22. <https://doi.org/10.1080/17512786.2019.1697956>.
- Kashyap G & Bhaskaran H (2020) Teaching data journalism: Insights for Indian journalism education. *Asia Pacific Media Educator* 30 (1):44-58. <https://doi.org/10.1177/1326365X20923200>.
- Lewis NP (2021) Defining and teaching data journalism: A typology. *Journalism & Mass Communication Educator* 76 (1):78-90. <https://doi.org/10.1177/1077695820924309>.
- Lewis NP & Waters S (2018) Data journalism and the challenge of the Shoe-Leather epistemologies. *Digital Journalism* 6 (6):719-736. <https://doi.org/10.1080/21670811.2017.1377093>.
- Li L (2022) Data news dissemination strategy for decision making using new media platform. *Soft Computing* 26: 10677-10685. <https://doi.org/10.1007/s00500-022-06819-0>.
- Morini F (2023) Data journalism as “Terra in Cognita”: Newcomers tensions in shifting towards data journalism epistemology. *Journalism Practice* 1-18. <https://doi.org/10.1080/17512786.2023.2185656>.

AR: "Trends in examining data journalism research"

- Neyazi TA & Schroeder R (2021) Was the 2019 Indian election won by digital media?. *The Communication Review* 24 (2):87-106. <https://doi.org/10.1080/10714421.2021.1933311>.
- Ojo A & Heravi B (2018) Patterns in award winning data storytelling: Story types, enabling tools and competencies. *Digital Journalism* 6 (6):693-718. <https://doi.org/10.1080/21670811.2017.1403291>.
- Parasie S (2022) *Computing the news: Data journalism and the search for objectivity*. New York: Columbia University Press.
- Pentzold C & Fechner D (2021) Probabilistic storytelling and temporal exigencies in predictive data journalism. *Digital Journalism* 9 (6):715-736. <https://doi.org/10.1080/21670811.2021.1878920>.
- Petticrew M & Roberts H (2006) *Systematic review in the social sciences*. Malden: Blackwell Publishing.
- Richards A (2023) *Foundations of data and digital journalism*. New York & London: Routledge.
- Schaetz N, Lischka JA, & Laugwitz L (2023) Datafication of journalism: How data elites and epistemic infrastructures change news organizations. *Digital Journalism* 1-19. <https://doi.org/10.1080/21670811.2023.2235601>.
- Shaffril HAM, Samsuddin SF, & Samah AA (2020) The ABC of systematic literature review: The basic methodological guidance for beginners. *Quality & Quantity* 55: 1319-1346. <https://doi.org/10.1007/s11135-020-01059-6>.
- Stalph F & Heravi B (2021) Exploring data visualisations: An analytical framework based on dimensional components of Data Artefacts in journalism. *Digital Journalism*:1-23. <https://doi.org/10.1080/21670811.2021.1957965>.
- Uretta AL & Rodriguez EMF (2021) The potential of investigative data journalism to reshape professional culture and value. A Study of bellwether transnational projects. *Communication & Society* 34 (1):41-56. <https://doi.org/10.15581/003.34.1.41-56>.
- Veglis A & Maniou TA (2018) The mediated data model of communication flow: Big data and data journalism. *KOME-An International Journal of Pure Communication Inquiry* 6 (2):32-43. <https://doi.org/10.17646/KOME.2018.23>.
- Vrabec N & Proner J (2021) Data journalism in the context of important political events: The case of Slovakia. *Medijske Studije* 12 (24):40-54. <https://doi.org/10.20901/ms.12.24.3>.
- Vural ZI & Masip P (2021) Data journalism as an innovation in social communication: The case in sports industry. *European Public & Social Innovation Review* 6 (1):42-55.
- Wang Q, Guant T, & Yan X (2023) The "Conditional Unbias" of data journalism in China's news industry. *Digital Journalism* 1-20. <https://doi.org/10.1080/21670811.2023.2278045>.
- Widholm A & Appelgren E (2022) A softer kind of hardnews? Data journalism and the digital renewal of public service news in Sweden. *New media & Society* 24 (6):1363-1381. <https://doi.org/10.1177/1461444820975411>.
- Wright, S & Nolan, D (2023) The development of data journalism in China: Influences, motivations, and practice. *Digital Journalism* 11 (9):1664-1681. <https://doi.org/10.1080/21670811.2021.1927779>.
- Wright K, Zamith R, & Bebawi S (2019) Data journalism beyond majority world countries: Challenges and opportunities. *Digital Journalism* 7 (9):1295-1302. <https://doi.org/10.1080/21670811.2019.1702472>.
- Wu S (2023) Data journalism and the impact of politics: Examining data stories produced across different political systems. *Journalism Practice*. <https://doi.org/10.1080/17512786.2023.2263753>.
- Wu S (2022) Evaluating "exemplary data journalism" from Asia: An exploration into South China Morning Post's data stories on China and the world. *Journalism* 1-17. <https://doi.org/10.1177/14648849221093509>.
- Wu S (2021) Data "objectivity" in a time of coronavirus: Uncovering the potential impact of state influence on the production of data-driven news. *Digital Journalism* 9 (9):1303-1320. <https://doi.org/10.1080/21670811.2021.1942111>.
- Young ML & Hermida A (2015) From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. *Digital Journalism* 3 (3):381-397. <https://doi.org/10.1080/21670811.2014.976409>.
- Zhang X & Chen M (2020) Journalists' adoption and media's coverage of data-driven journalism: A case of Hong Kong. *Journalism Practice*. <https://doi.org/10.1080/17512786.2020.1824126>.

Author Biography

M Fikri AR is a Ph.D. student at the Department of Communication, Institute of Social Sciences Sakarya University, Turkey. He is a recipient of the YTB Scholarship (Türkiye Bursları), with research interests in Data Journalism, Politics, and Media. He is also a lecturer at the Department of Communication Science, Faculty of Social and Political Sciences, Brawijaya University, Indonesia.