

Open data journalism

a domain mapping review

Papageorgiou, Georgios; Euripides, Loukis; Magnussen, Rikke; Charalabidis, Yannis

Published in:

Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance, ICEGOV 2023

DOI (link to publication from Publisher):

[10.1145/3614321.3614340](https://doi.org/10.1145/3614321.3614340)

Creative Commons License

CC BY 4.0

Publication date:

2023

Document Version

Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Papageorgiou, G., Euripides, L., Magnussen, R., & Charalabidis, Y. (2023). Open data journalism: a domain mapping review. In D. Getschko, I. Lindgren, & M. Yildiz (Eds.), *Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance, ICEGOV 2023* (pp. 159-166). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3614321.3614340>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



OPEN DATA JOURNALISM: A DOMAIN MAPPING REVIEW

Georgios Papageorgiou*

Department of Information and Communication Systems
Engineering, University of the Aegean, Greece
gpapag@aegean.gr

Rikke Magnussen

ICT and Design for Learning, Department of
Communication, Aalborg University, Copenhagen,
Denmark
rikkem@hum.aau.dk

Euripides Loukis

Department of Information and Communication Systems
Engineering, University of the Aegean, Greece
eloukis@aegean.gr

Yannis Charalabidis

Department of Information and Communication Systems
Engineering, University of the Aegean, Greece
yannisx@aegean.gr

ABSTRACT

Journalism is highly important in modern societies, both for shaping public opinion and also as gatekeeper of transparency and accountability, being considered as ‘the fourth constitutional power’. The increasing dissemination of digital technologies and data had a profound impact on journalism, giving rise to the development of ‘data journalism’. The advent of open data provides great opportunities to journalism, towards gaining a better insight into important social problems and government activities, and based on them develop high quality ‘evidence-based’ journalistic articles and work in general, giving rise to the development of ‘open data journalism’. Even though this is of great importance for our societies, and journalists are a very special and important user group in open data research, there is not much research literature about it. This paper conducts a systematic literature review on open data journalism scientific research, and provides a ‘map’ of this important domain, which includes the main research themes that have been investigated; based on them three main future research directions have been formulated, and for each of them appropriate theoretical foundations are proposed. The findings of this study reveal as a main issue the need for new competencies concerning digital technologies as well as data analysis that journalists will have to develop, as well as the necessity for co-operation with other actors for a higher exploitation and more sophisticated analysis of open data, which might result in the development of open data journalism ecosystems.

CCS CONCEPTS

• **Applied Computing**; • **Computers in other domains**; • **Computing in Government**; • **E-government**; • **Applied computing**; • **General and reference**; • **Document types**; • **Surveys and overviews**;

*Corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICEGOV 2023, September 26–29, 2023, Belo Horizonte, Brazil
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0742-1/23/09.
<https://doi.org/10.1145/3614321.3614340>

KEYWORDS

Open Data, Data Journalism, Journalism, Systematic Literature Review

ACM Reference Format:

Georgios Papageorgiou, Euripides Loukis, Rikke Magnussen, and Yannis Charalabidis. 2023. OPEN DATA JOURNALISM: A DOMAIN MAPPING REVIEW. In *16th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2023)*, September 26–29, 2023, Belo Horizonte, Brazil. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3614321.3614340>

1 INTRODUCTION

Journalism is highly important in modern societies, both for shaping public opinion and also as gatekeeper of transparency and accountability, being considered as ‘the fourth constitutional power’ [1]. The increasing dissemination of digital technologies and data had a profound impact on journalism, giving rise to the development of ‘data journalism’ [2]. The importance of data in the journalistic process is pivotal as they enable journalists with the ability to discover and support their publications with unshakable evidence. By having solid evidence in their hands, they can be more effective in monitoring the government and other institutions and elevate their role as guardians of transparency and accountability in democratic societies. Data journalism has emerged as a new evolution of journalism in recent years; this new form of journalism is defined by Veglis and Bratsas (2017) [1] as the use of data in all the stages of the journalistic process, the extraction of information, the compilation and the visualisation in a comprehensive way. And as experts have reported, this new form of journalism has increased in popularity the recent years [2].

The first known use of data for journalistic purposes dates back to the 19th century when leaked data set in the form of a table for the state of schools in Manchester was published in the Manchester Guardian newspaper in 1821 [3]. Also, in 1859, the first use of infographics in a news article about the mortality rate from infectious diseases in the British army during the Crimean War [4] was introduced. But the use of computers for journalistic purposes will still want a century to emerge, as computers were initially used in 1952 to forecast the results of the USA presidential election [5]. This use of computers in the journalistic process was named Computer-assisted reporting (CAR). Although this innovative attempt was successful, this new technology will take one more decade to be widely adopted. The next innovative step is attributed to Philip

Meyer, a Detroit Free Press journalist who used computers to analyse the data from the riots in Detroit during the late 1960s [6]. But it was during the 2000s that the use of computers integrated into the working environments, and this led to the emergence of a new term: Data Journalism. Data journalism was quoted for the first time by Simon Rogers in 2008 in the Guardian Insider Blog [7]. While CAR is viewed as a technique for gathering and analysing data to enhance investigative journalism, data journalism refers to using data throughout the journalistic process [6]. Another difference between them was the availability of data; while in the past, journalists were more focused on the collection of data and the creation of their own data sources, in the age of data journalism, the high availability of data shifted the focus the collection to the analysis and the presentation.

The abundance of data we experience today, which contributed to the emergence of data journalism, can be attributed, in part, to the rise of open data initiatives [8]. Open data is defined as “Data that is available free of charge, openly licensed, and in an open, machine-readable format” [9], and it was quoted for the first time in the publication “On the Full and Open Exchange of Scientific Data” in 1995 [10]. Although the movement of open data will need almost another decade from its first mention to get greater recognition, two were the important milestones for the wide adoption of open data. The Open Knowledge Foundation (<https://okfn.org>) that launched in 2004, and 2009 the launch of Data.gov by the USA government. After that, other countries joined the open data movement and started to release their data to increase innovation and transparency. This progression significantly increased the available data sources journalists could access [11].

Open data and journalism have been well-explored separately in academia, but there is a research gap concerning their combination: ‘open data journalism’ meant as data journalism using open data.

This lack of research on the intersection of these two topics is surprising since journalism and open data have several commonalities, making information more accessible to the public and enforcing transparency. Also, the increasing availability of data in the past few decades and their practical use from several big news organisations make the absence of research on open data journalism even more peculiar. As open data policies continue to be adopted by governments worldwide [12], journalism will have the opportunity benefit greatly by using them, and become an even more critical pillar for transparency and accountability in the years to come. Therefore, studying the joint research concerning open data and journalism is essential since these two topics create a synergy that can improve democratic function. This paper aims to support and facilitate this research, by conducting a systematic literature review (SLR) [13], [14] of open data journalism scientific research, and providing a ‘map’ of this important domain, which includes the main research themes that have been investigated. So, our research question is:

RQ: “What are the main themes under examination in the published papers on open data journalism?”

Based on the identified research themes existing research gaps are determined and future research directions are proposed as well as appropriate theoretical foundations for each of them.

Our paper consists of five sections. The following section, 2, describes the methodology of our literature review. In section 3 the results are presented, and then discussed in section 4, while in the final section 5 the conclusions are summarised, and future research directions are proposed.

2 METHODOLOGY

The literature review method used for the analysis of the previous literature on open data journalism and the identification of the main themes is a systematic literature review [13], [14] (SLR); so our

Table 1: Inclusion/exclusion criteria

Inclusion/exclusion criteria	Description: Records selected for analysis should fulfil the following criteria:
Open Data	The publication must have a focus on the use of open data.
Journalism	The publication must focus on journalism
Publication quality	The documents must peer-reviewed.
Language	The publications must be in the English language.
Time frame	The publications must be published between 2010 and 2022

Table 2: Search queries

Data Base	Queries
IEEE	("open data" OR "open government data") AND ("Journalism" OR "Journalists" OR "Journalist")
Web of Science	(("open data") OR ("open government data") AND ("Journalism") OR ("Journalists") OR ("Journalist")) (Title) or (("open data") OR ("open government data") AND ("Journalism") OR ("Journalists") OR ("Journalist")) (Abstract) or (("open data") OR ("open government data") AND ("Journalism") OR ("Journalists") OR ("Journalist")) (Keyword Plus ®)
Scopus	(TITLE-ABS-KEY ("open data") OR TITLE-ABS-KEY ("open government data") AND TITLE-ABS-KEY ("Journalism") OR TITLE-ABS-KEY ("Journalists") OR TITLE-ABS-KEY ("Journalist"))
Science Direct	(("open data") OR ("open government data")) AND (("Journalism") OR ("Journalists") OR ("Journalist"))

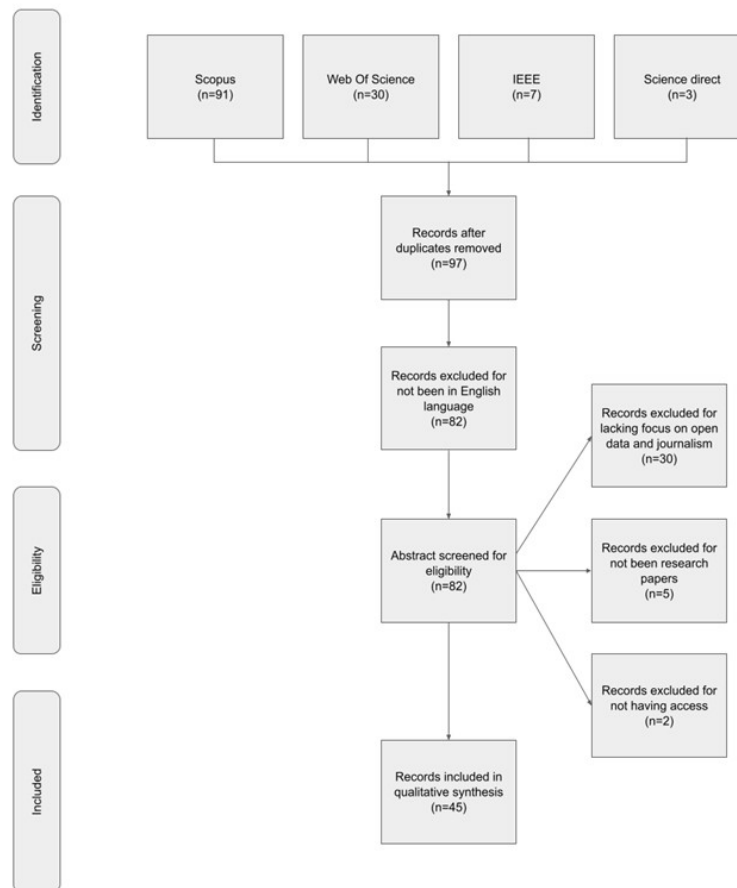


Figure 1: Systematic literature review process.

review process includes formulating the research question, defining the inclusion and exclusion criteria for the literature search, and then searching for publications with a formal process with the use of strictly defined keywords and screening the papers according to the defined inclusion and exclusion criteria.

2.1 Defining the Inclusion and Exclusion Criteria

The inclusion and exclusion criteria aim to ensure that only publications relevant to the research topic will be found and analysed; we can see them in Table 1 the publication must have a focus on open data as well as on journalism. The publications must be peer-reviewed to ensure our findings’ quality; they also have to be in English so that we can understand their content. Finally, we selected a timeframe for the publications (2010-2022).

2.2 Discovering the Literature and Applying the Screening Process

The selected keywords for our literature search correspond to our two main focuses, journalism and open data, and are shown in Table 2 with respect to the former “journalism”, “journalists”, and

“journalist” have been used (other keywords, such as “media” or “reporting,” were deemed unsuitable since the results were not related only to journalism); with respect to the latter the term “open government data” was also used along with “open data” (since the scientific papers databases we used the databases returned only exact matches, when using the “open data” keyword, all the publications with the term “open government data” were not included).

Using the above queries to these four the databases 131 publications were extracted. The next step was to remove the duplicates, and filter out the ones that were not in English, and we ended up with 82 publications (excluding 34 duplicates, and also 11 publications in Spanish, 2 in Portuguese, 1 in Turkish and 1 in French). Based on the titles and abstracts of the papers, we conducted the screening processes and applied the rest of the exclusion criteria to these 82 documents; we ended up with 45 ones. The details of the whole process we followed are presented in Figure 1.

From these papers that met the inclusion criteria mentioned, which are shown in the second column of Table 3, we extracted the research themes they investigated using an open-coding approach [15]. In particular, for each paper, we read the title and the abstract (and, if required, some parts of its text) to understand the central theme it investigates, and then we generalise it; next, we compare

it with the previously identified themes, and if it belongs to one of them, we associate it with this theme, while if it is new, we express it in an abstract manner concisely using 3-5 words and add it to the list of the identified themes.

3 RESULTS

Using the methodology described in the previous section nine research themes were identified, which are shown in the following Table 3, and analysed in sub-sections 3.1 – 3.9.

3.1 Tools for Journalists

The publications of this theme describe software tools that have been developed for journalists for enhancing their work using open data. Two main categories were detected, the tools that are focused on data analysis and support the journalist’s process in its entirety and the tools that provide the ability to discover and evaluate news instantly. Most of the publications belonged to the first category, describing software tools that cover all the parts of the journalistic process, such as data collection, data analysis, visualisation and presentation.

An interesting study that focuses on visualisation concerns the case of the Indian elections [31]. The goal of the tool that was developed and presented was to facilitate the display of voting in India in a way that could be easily understandable by the public. The main problems they faced were the immense volume of votes they had to display and, secondly, to provide a way to navigate it. Concerning another aspect of the journalistic process, we have publications focusing on tools designed to make the data analysis easier. An example is The Gamma [29], a low coding tool journalists can use to analyse public data without prior programming knowledge. Also, concerning the discoverability aspect, we have publications presenting tools like GovWild [26] that can integrate various data sources and provide the functionality of searching all of them. In particular, GovWill can combine and clean open government data sets and then produce one linked open dataset. On the other category of tools, the ones that are designed to discover news, we have the case described by Gotttron et al. (2015) [24], where a system is presented that can help the journalist to find geospatial-related data in real time.

3.2 Impact of Open Data on Journalism

Only three papers focused on the impact of digital technologies on journalism. One paper focuses on the use of big data [34] and how this technology will change journalism, since it will provide the ability to access and process great volumes of information and, therefore, providing new deep insights into important social problems and government activities. Another paper focuses on interactive maps [35] and their potential to visualise complicated data topics comprehensively. Finally, we have one publication on the impact of Artificial intelligence in journalism [36]; in this publication, a variety of potential uses are presented, like fake news detection and voice-to-text software.

3.3 Open Data Journalism Practices

The theme of Journalism Practises contains publications focusing on practices of open data use in the day-to-day work of journalists. Almost all the papers on this theme mention the journalist’s problem of lack of technical skills. An interesting finding related to that problem is that this lack of skills works as a barrier to the exploitation of the open data that governments publish and discourages journalists from getting involved with data-related activities [37]. Another widely encountered problem mentioned is the availability of open data sources. In the examination of the state of data journalism in Italy, is mentioned the struggle to find data, and in many cases, the data they get are of bad quality [38]. Finally, another interesting finding was in the examination of articles published during Women’s Day in Brazil in 2017 [39]. This publication focused on analysing all the published articles of that year from three major Brazilian newspapers to evaluate the use of data in journalism. The findings indicated that the journalists in most of the publications didn’t provide data sources or mention where they found their infographics.

3.4 Journalists’ Data Literacy

Four papers focused on journalists’ skills and abilities as well as their education concerning the use of data, especially open data. In this theme, two papers focus on formal education. The first focuses on comparing educational programs on data journalism in European countries [42], and the second focuses on the reforms that must be made in the curricula for data analysis to accommodate the increase in data availability we are experiencing [43]. An interesting publication was “On Some Russian Educational Projects in Open

Table 3: Thematic classification of the selected publications

Theme	Papers	No. studies
Software Tools for Journalists	[16–33]	18
Impact of Open Data on Journalism	[34, 34–36]	3
Open Data Journalism Practices	[37–40]	4
Journalists Data Literacy	[41–44]	4
Collaboration	[45–50]	6
Legislation and Ethics	[51–53]	3
Cases of Open Data Use for Journalism	[54–57]	4
Communication Methods	[58, 59]	2
Software Tools for the Public	[60]	1

Data and Data Journalism” [44], where a series of data journalism workshops were conducted in Russia. It also mentions that the workshops were quite introductory in data journalism. It advocates that the workshops only cover introductory material to the subject and that more specialised workshops must be organised. Finally, we have a publication that is not dealing with formal education; its main argument is that technical skills are not enough, but also experience in statistics is required for someone to be open data literate [41].

3.5 Collaboration

The papers categorised in the collaboration theme can be divided into two categories: collaboration with the public and collaboration with other professionals. In the former category concerning collaboration with the public, the main argument in the publications was that this could promote accountability and transparency, since the citizens are directly involved in the data analysis as they can have more experience than journalists in specialised topics. In [47] is analysed an investigative journalism story run by The Guardian that combined open data, crowdsourcing and game mechanics with the purpose of engaging readers. In the latter category concerning collaboration with experts, the main focus is to address the data literacy problems of the journalists. The papers analyse several cases of collaboration of journalists with different professions and ways to find and make experts outside of journalism interested in data analysis and visualisation for journalistic purposes. Interesting cases are the use of hackathons to find experts in the technology sector [49] and the involvement of civil activists and open data hackers [50].

3.6 Legislation and Ethics

Two of the publications categorised in this theme focus on legislation comparison. The publication “The Impact of Public Transparency Infrastructure on Data Journalism” [52] presents a comparative study between counties with respect to the transparency level they achieve through different legislation for data accessibility; also, there is comparison between the Right for Information Act and the Open Data government. The research examines to what extent these acts are applied and their impact on the data available to journalists. The focus of the publication “The Right to Know Through the Freedom of Information and Open Data” [53] is a comparative analysis between the Freedom of information act and the Open Data regulations in five countries. Finally, we have a publication that focuses on the ethical issues that can arise when using open data [51], especially focusing on the potential harm from the possibility of de-anonymising data and exposing people’s identities.

3.7 Cases of Open Data Use for Journalism

The four papers of this theme examine cases of journalistic articles that use open data in several counties, but the focus of each is different. In [56] the focus is on how the news media used data in their reporting during the NHS (National Health Service, UK) winter crisis of 2016–2017. Another interesting focus is the case of the open parliamentary data in Norway, Sweden and Denmark [55], where the journalist and other actors using open data are struggling to find appropriate valuable datasets; it also mentions the problem

of fragmentation of data across various governmental agencies and the need for national open data repositories.

3.8 Communication Methods

Two papers were categorised in this theme. The first paper [59] uses the ‘mediated data model of communication flow’ to shed light on the current communication process between journalists/media and their initial sources of digital information, using big data as case study. The second paper focuses on the different visualisation techniques that can be applied to present and communicate complex and linked datasets [58]. The difference between this publication and the others examining visualization techniques is that this focuses on the methodologies rather than technical specifications.

3.9 Software Tools for the Public

Only one paper was categorised in this theme [60]. It presents a software tool that can be used by citizens for accessing open data about the legislation in Brazil concerning their political preferences and interests, and avoiding the filtering performed quite often by the journalists based on media agenda. It argues that journalists cannot be the gatekeepers of information as they can be prone to biases, and therefore the public has to bypass them and have direct access to data.

4 DISCUSSION

By examining the nine research themes we identified in this open data journalism domain, and the number of papers of each theme, which are shown in Table 3, the first conclusion that can be drawn is that this research domain is still in its infancy, as more than half of the 45 publications on open data journalism we found are either descriptions of software tools to be used by journalists (18 publications of the largest thematic category on ‘Software Tools for Journalists’) and also by simple citizens for similar purposes (1 publication of the ‘Software Tools for the Public’ thematic category), or relevant case studies (4 publications of the ‘Cases of Open Data Use for Journalism’ thematic category). Much smaller is the number of papers concerning the ‘core’ of this domain: the ways and practices of open data use by journalists (4 publications of the ‘Open Data Journalism Practices’ thematic domain), as well as the impact of the open data on journalism (3 publications of the ‘Impact of Open Data on Journalism’ thematic category). Furthermore, most of the publications we found on open data journalism are based on case studies or interviews with small numbers of journalists (or other stakeholders) but are missing larger scale surveys of large numbers of participants, which would provide more generalizable conclusions; also, most of the publications lacked sound theoretical foundations.

An important share of the open data journalism publications we found (the 4 publications of the ‘Journalists Data Literacy’ thematic category, and the 6 publications of the ‘Collaboration’ one) are dealing with an inherent ‘structural’ problem of open data journalism: in order to find and use open data effectively and extract deep insights and knowledge from them journalists need substantial technological as well as statistical data processing skills and abilities, which the vast majority of them do not have, and this constitutes a big barrier to open data exploitation by journalists; so on one hand

relevant education of journalists is required, while on the other hand journalists have to co-operate extensively with technological and statistical experts, and this might result in the development of open data journalism ecosystems. It should be noted that in many of the publications of the other thematic categories as well we identified mentions of the technical skills that journalists need in order to find and use open data properly; this was more prominent in the ‘Open Data Journalism Practices’ and the ‘Cases of Open Data Use for Journalism’ thematic categories. Also, in many publications are proposed possible solutions to this problem: in publications of the ‘Software Tools for Journalists’ and ‘Collaboration’ thematic categories we mostly encountered ways journalists can use open data without acquiring deep technical knowledge. This can be achieved using easy to use tools that can help in the open data search, collection, analysis and visualisation; furthermore, in the ‘Collaboration’ thematic category, the overwhelmingly mentioned solution was to work closely with other experts that can cover the skill gaps the journalists have.

From a more detailed examination of the publications we have also found it can be concluded that the themes of ‘Open Data Journalism Practices’ and ‘Cases of Open Data Use for Journalism’ can be considered as ‘different sides of the same coin’ since they focus on the adoption of open data in journalism, but they also have significant differences. Most publications in the ‘Open Data Journalism Practices’ theme focus on journalists as individuals, and the research methodologies they use are primarily interviews and secondarily surveys. On the other hand, the publications on the ‘Cases of Open Data Use for Journalism’ themes are mainly case studies, and they are not focused on journalists but on the result of their work using open data. However, the number of publications this category is quite limited (only 4); more case studies are deemed quite necessary, since they allow gaining a better insight about not only the adoption of open data in the journalistic profession but also the impact that they can have on the profession and the society in general (to what extent the use of open data by journalists can lead to a higher quality of ‘evidence-based’ journalism, and also to what extent the use of open data or/and conclusions drawn from them in journalistic articles can make them more influential to the public).

Another interesting finding from the exclusion process was that 17% of the papers we discovered were not in the English language; we consider that to be quite remarkable since English is the predominant language in academic publications; 12% of these papers were in Spanish and 2% in Portuguese, published mostly from counties in South America. We do not have proficient knowledge of Spanish, so we could not read these publications, but their volume made us formulate the assumption that there is a separate ecosystem in South America that is researching the topic of open data journalism, and this can be another source of research inspiration.

Finally, based on the above review and mapping of open data journalism scientific research, and the research gaps we have identified, we can formulate three main directions of future research in this domain, together with appropriate theoretical foundations for each of them, which can contribute to increasing its maturity; they include qualitative studies (based on in-depth interviews and focus-groups) as well as quantitative studies (based on large scale surveys) on:

RD1: Perceptions of journalists and other stakeholders about various aspects of open data, such as ease of use and usefulness, based on the ‘Technology Acceptance Model’ (TAM) [61-63], or comparative advantage, complexity, compatibility, trialability and observability, based on the ‘Diffusion of Innovation’ (DOI) theory [64-66], or performance expectancy, effort expectancy, facilitating conditions, social influence and hedonic motivation, based on the Unified Theory of Acceptance and Use of Technology (UTAUT) [67]; and also attitudes and intentions towards the use of open data for journalistic purposes, as well as real use (if it exists).

RD2: Ways and forms of open data use for various kinds of journalistic work, including investigative journalism, as well as barriers and difficulties; in general investigation of the ‘positive affordances’ and the possible ‘negative affordances’ of the use of open data in journalism, both the ‘perceived’ and the ‘actualized’ affordances, using the lenses of ‘affordances theory’ [68-70] and exploiting the methods and knowledge in general developed in this area.

RD3: Benefits and impact of open data use for various kinds of journalistic work, concerning its efficiency, effectiveness and innovation, at various levels: individual, organizational (examining various types of journalism organizations, such as newspapers, news portals, etc.) and sectoral (i.e. for journalism in general); for these purposes a useful theoretical foundation can be the DeLone and McLean’s Information Systems Success Model [71-73].

5 CONCLUSIONS

In this paper, we used the structured literature review methodology [13-14] in order to review the academic literature on the intersection of journalism with open data, or open data journalism, and develop a thematic map of this domain. We searched four scientific databases, and after a process of excluding some papers based on pre-defined criteria, we ended up with 45 publications, which were analysed and grouped into thematic categories in order to determine the main research themes of the open data journalism domain. Nine research themes were identified, with the largest of them (i.e. the one with the highest number of publications) being ‘Software Tools for Journalists’ as well as ‘Collaboration’ (between journalists and technological and statistical experts). By examining the above publications on open data journalism, as well as the above research themes we identified, it can be concluded that this research domain is still in its infancy, as more than half of the 45 publications on open data journalism we found are either descriptions of relevant software tools to be used by journalists or simple citizens, or relevant case studies. Furthermore, research gaps have been identified, based on them three main future research directions have been formulated, and for each research direction appropriate theoretical foundations have been proposed.

The thematic map of the open data journalism domain developed in this study can be very useful for gaining an overall picture of this domain, as well as for supporting and facilitating the extensive future research required on this topic, while the identified research gaps and the proposed future research directions can be very useful for orienting this research; furthermore, the theoretical foundations we have proposed can enhance the quality, the completeness, the reliability and finally the usefulness of this research. As mentioned

in previous sections the first step in our research endeavour was to discover if there are other literature reviews on open data journalism, and, to the best of our knowledge, we have not discovered any; therefore, our work can be used as a guide for other researchers and professionals involved in open data journalism.

Further research is required in order to gain a better understanding of the timewise evolution of open data journalism research, the countries in which it has been conducted, the research approaches and methods it has used, and also its theoretical foundations (for the limited number of publications having some theoretical foundation). Finally, it is necessary to investigate further the barriers and benefits journalists encounter when using open data that have been identified by the literature, as well as the proposed ways and actions for addressing the former and increasing the latter.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569.

The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

- [1] A. Veglis and C. Bratsas, "Reporters in the age of data journalism," *J. Appl. Journal. MEDIA Stud.*, vol. 6, no. 2, pp. 225–244, Jun. 2017, doi: 10.1386/ajms.6.2.225_1.
- [2] M. Knight, "Data journalism in the UK: a preliminary analysis of form and content," *J. Media Pract.*, vol. 16, no. 1, pp. 55–72, Jan. 2015, doi: 10.1080/14682753.2015.1015801.
- [3] S. Rogers, "The first Guardian data journalism: May 5, 1821," *the Guardian*, Sep. 26, 2011. <http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian> (accessed Apr. 17, 2023).
- [4] S. Rogers, "Florence Nightingale, datajournalist: information has always been beautiful," *The Guardian*, Aug. 13, 2010. Accessed: Apr. 17, 2023. [Online]. Available: <https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics>
- [5] B. Houston, *Computer-assisted reporting: a practical guide*, Fourth Edition. New York; London: Routledge, Taylor & Francis Group, 2015.
- [6] J. Gray, L. Bounegru, L. Chambers, European Journalism Centre, and Open Knowledge Foundation, Eds., *The data journalism handbook*, 1st ed. Sebastopol, CA: O'Reilly Media, 2012.
- [7] S. Rogers, "Turning official figures into understandable graphics, at the press of a button," *The Guardian*, Dec. 18, 2008. Accessed: Apr. 18, 2023. [Online]. Available: <https://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata>
- [8] Y. Charalabidis, C. Alexopoulos, E. Ferro, M. Janssen, T. Lampoltshammer, and A. Zuiderwijk, *The World of Open Data: Concepts, Methods, Tools and Experiences*, 1st ed. 2018. in Public Administration and Information Technology, no. 28. Cham: Springer International Publishing; Imprint: Springer, 2018. doi: 10.1007/978-3-319-90850-2.
- [9] B. Van Loenen *et al.*, "Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda," *JeDEM - EJournal EDemocracy Open Gov.*, vol. 13, no. 2, pp. 1–27, Dec. 2021, doi: 10.29379/jedem.v13i2.644.
- [10] *On the Full and Open Exchange of Scientific Data*. Washington, D.C.: National Academies Press, 1995. doi: 10.17226/18769.
- [11] T. Davies, S. B. Walker, M. Rubinstein, and F. Perini, "The State of Open Data: Histories and Horizons".
- [12] Publications Office of the European Union., *Open data maturity report 2021*. LU: Publications Office, 2022. Accessed: May 02, 2023. [Online]. Available: <https://data.europa.eu/doi/10.2830/394148>
- [13] C. Okoli, "A Guide to Conducting a Standalone Systematic Literature Review," *Commun. Assoc. Inf. Syst.*, vol. 37, 2015, doi: 10.17705/1CAIS.03743.
- [14] Y. Xiao and M. Watson, "Guidance on Conducting a Systematic Literature Review," *J. Plan. Educ. Res.*, vol. 39, no. 1, pp. 93–112, Mar. 2019, doi: 10.1177/0739456X17723971.
- [15] H. Maylor, K. L. Blackmon, and M. Huemann, *Researching business and management*, 2nd edition. London: Palgrave, 2017.
- [16] T.-D. Cao, L. Duroyon, F. Goasdoué, I. Manolescu, and X. Tannier, "BeLink: Querying networks of facts, statements and beliefs," in *Int Conf Inf Knowledge Manage*, Association for Computing Machinery, 2019, pp. 2941–2944. doi: 10.1145/3357384.3357851.
- [17] M. G. Ocaña and A. L. Opdahl, "Challenges and opportunities for journalistic knowledge platforms," *CEUR Workshop Proceedings*, vol. 2699. CEUR-WS, 2020. [Online]. Available: [https://www.scopus.com/inward/record.uri?eid=\\$-s2.0-85097522097&partnerID=\\$40&md5=\\$90baef640c01a42989fd0237bfb401ec](https://www.scopus.com/inward/record.uri?eid=$-s2.0-85097522097&partnerID=$40&md5=$90baef640c01a42989fd0237bfb401ec)
- [18] S. Bozsik, X. Cheng, M. Kuncham, and E. Mitchell, "Democratizing Housing Affordability Data: Open Data and Data Journalism in Charlottesville, VA," in *2022 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA: IEEE, Apr. 2022, pp. 178–183. doi: 10.1109/SIEDS55548.2022.9799410.
- [19] M. G. Ocaña, A. L. Opdahl, and D.-T. Dang-Nguyen, "Emerging News task: Detecting emerging events from social media and news feeds," *CEUR Workshop Proceedings*, vol. 3181. CEUR-WS, 2021. [Online]. Available: [https://www.scopus.com/inward/record.uri?eid=\\$-s2.0-85137038782&partnerID=\\$40&md5=\\$45f5a4bfaeb6ec928f378748892f71fa](https://www.scopus.com/inward/record.uri?eid=$-s2.0-85137038782&partnerID=$40&md5=$45f5a4bfaeb6ec928f378748892f71fa)
- [20] V. Shehu, A. Mijushkovic, and A. Besimi, "Empowering data driven journalism in Macedonia," *ACM International Conference Proceeding Series*. Association for Computing Machinery, 2016. doi: 10.1145/2955129.2955187.
- [21] A. Rind, D. Pfahler, C. Niederer, and W. Aigner, "Exploring media transparency with multiple views," *CEUR Workshop Proceedings*, vol. 1734. CEUR-WS, pp. 65–73, 2016. [Online]. Available: [https://www.scopus.com/inward/record.uri?eid=\\$-s2.0-84999788893&partnerID=\\$40&md5=\\$36d6f68c7b94403102106e6b8b563ea](https://www.scopus.com/inward/record.uri?eid=$-s2.0-84999788893&partnerID=$40&md5=$36d6f68c7b94403102106e6b8b563ea)
- [22] T. D. Cao, I. Manolescu, and X. Tannier, "Extracting Linked data from statistic spreadsheets," in *Proc. Int. Workshop Semant. Big Data, SBD - conjunction ACM SIGMOD/PODS Conf.*, Gruenwald L. and Groppe S., Eds., Association for Computing Machinery, Inc, 2017. doi: 10.1145/3066911.3066914.
- [23] F. Goasdoué, J. Leblay, K. Karanasos, I. Manolescu, Y. Katsis, and S. Zampetakis, "Fact checking and analyzing the Web," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, New York, NY, 2013, pp. 997–1000. doi: 10.1145/2463676.2463692.
- [24] T. Gotttron, J. Schmitz, and S. E. Middleton, "Focused exploration of geospatial context on Linked Open Data," *CEUR Workshop Proceedings*, vol. 1279. CEUR-WS, 2015. [Online]. Available: [https://www.scopus.com/inward/record.uri?eid=\\$-s2.0-84939629914&partnerID=\\$40&md5=\\$ad78caf3e814bd52dba399a2f71b2456](https://www.scopus.com/inward/record.uri?eid=$-s2.0-84939629914&partnerID=$40&md5=$ad78caf3e814bd52dba399a2f71b2456)
- [25] A. Paley *et al.*, "From data to information: Automating data science to explore the U.S. court system," *Proceedings of the 18th International Conference on Artificial Intelligence and Law, ICAIL 2021*. Association for Computing Machinery, Inc, pp. 119–128, 2021. doi: 10.1145/3462757.3466100.
- [26] C. Böhm *et al.*, "Linking open government data: What journalists wish they had known," *ACM International Conference Proceeding Series*. 2010. doi: 10.1145/1839707.1839751.
- [27] Y.-A. Le Borgne, A. Homolova, and G. Bontempi, "OpenTED browser: Insights into European Public Spendings," in *CEUR Workshop Proc.*, Gama J., Zliobaite I., Zliobaite I., and Gavalda R., Eds., CEUR-WS, 2016. Accessed: Sep. 19, 2016. [Online]. Available: [https://www.scopus.com/inward/record.uri?eid=\\$-s2.0-85019195878&partnerID=\\$40&md5=\\$fa04428595f7474afce7eac77d25e760](https://www.scopus.com/inward/record.uri?eid=$-s2.0-85019195878&partnerID=$40&md5=$fa04428595f7474afce7eac77d25e760)
- [28] J. Klimek, J. Kučera, M. Nečaský, and D. Chlapek, "Publication and usage of official Czech pension statistics Linked Open Data," *Journal of Web Semantics*, vol. 48. Elsevier B.V., pp. 1–21, 2018. doi: 10.1016/j.websem.2017.09.002.
- [29] T. Petricek, "The Gamma: Programmatic Data Exploration for Non-programmers," in *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Roma, Italy: IEEE, 2022, pp. 1–7. doi: 10.1109/VL/HCC53370.2022.9833134.
- [30] F. J. Ekaputra *et al.*, "Towards open data mashups for data journalism," in *CEUR Workshop Proc.*, Hellmann S. and Fernandez J.D., Eds., CEUR-WS, 2018. Accessed: Sep. 11, 2017. [Online]. Available: [https://www.scopus.com/inward/record.uri?eid=\\$-s2.0-85045945245&partnerID=\\$40&md5=\\$9c60bb33b257a8dced67e405f170cadd](https://www.scopus.com/inward/record.uri?eid=$-s2.0-85045945245&partnerID=$40&md5=$9c60bb33b257a8dced67e405f170cadd)
- [31] K. Gupta, S. Sampat, M. Sharma, and V. Rajamanickam, "Visualization of election data: Using interaction design and visual discovery for communicating complex insights," *EJournal EDemocracy Open Gov.*, vol. 8, no. 2, pp. 59–86, 2016, doi: 10.29379/jedem.v8i2.422.
- [32] F. Evéquoz and H. Castanheiro, "Which Lobby Won the Vote? Visualizing Influence of Interest Groups in Swiss Parliament," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11685 LNCS. Springer, pp. 155–167, 2019. doi: 10.1007/978-3-030-27325-5_12.
- [33] N. Ó. Brocháin, L. Porwol, A. Ojo, T. Wagner, E. T. Lopez, and E. Karstens, "Extending open data platforms with storytelling features," *ACM International Conference Proceeding Series*, vol. Part F128275. Association for Computing Machinery, pp. 48–53, 2017. doi: 10.1145/3085228.3085283.
- [34] T. Sandoval-Martín and L. La-Rosa, "Big data as a differentiating sociocultural element of data journalism: The perception of data journalists and experts," *Commun. Soc.*, vol. 31, no. 4, pp. 193–209, 2018, doi: 10.15581/003.31.4.193-209.
- [35] D. A. Smith, "Online interactive thematic mapping: Applications and techniques for socio-economic research," *Comput. Environ. Urban Syst.*, vol. 57, pp. 106–117, 2016, doi: 10.1016/j.compenurbysys.2016.01.002.

- [36] A. Hassan and A. Albayari, *The Usage of Artificial Intelligence in Journalism*, vol. 1037. In *Studies in Computational Intelligence*, vol. 1037. Springer Science and Business Media Deutschland GmbH, 2022, p. 197. doi: 10.1007/978-3-030-99000-8_10.
- [37] C. Tabary, A.-M. Provost, and A. Trottier, "Data journalism's actors, practices and skills: A case study from Quebec," *Journalism*, vol. 17, no. 1, pp. 66–84, Jan. 2016, doi: 10.1177/1464884915593245.
- [38] C. Porlezza and S. Splendore, "From Open Journalism to Closed Data: Data Journalism in Italy," *Digit. J. Journal.*, vol. 7, no. 9, pp. 1230–1252, 2019, doi: 10.1080/21670811.2019.1657778.
- [39] A. C. Araújo, "Gender and Numbers: Using data from International Women's Day coverage on the sites of three major Brazilian newspapers," *Brazilian Journalism Research*, vol. 15, no. 1. Associação Brasileira de Pesquisadores de Jornalismo, pp. 74–101, 2019. doi: 10.25200/BJR.v15n1.2019.1061.
- [40] I. Radchenko and A. Sakoyan, *The view on open data and data journalism: Cases, educational resources and current trends*, vol. 436. In *Communications in Computer and Information Science*, vol. 436. Springer Verlag, 2014, p. 54. doi: 10.1007/978-3-319-12580-0_4.
- [41] J. Gray, C. Gerlitz, and L. Bounegru, "Data infrastructure literacy," *Big Data Soc.*, vol. 5, no. 2, 2018, doi: 10.1177/2053951718786316.
- [42] S. Splendore, P. Di Salvo, T. Eberwein, H. Groenhardt, M. Kus, and C. Porlezza, "Educational strategies in data journalism: A comparative study of six European countries," *Journalism*, vol. 17, no. 1, pp. 138–152, 2016, doi: 10.1177/1464884915612683.
- [43] J. Ridgway, "Implications of the Data Revolution for Statistics Education," *International Statistical Review*, vol. 84, no. 3. International Statistical Institute, pp. 528–549, 2016. doi: 10.1111/insr.12110.
- [44] I. Radchenko and A. Sakoyan, *On some russian educational projects in open data and data journalism*, vol. 9500. In *International Conference on Open Data for Education*, 2014, vol. 9500. Springer Verlag, 2016, p. 165. doi: 10.1007/978-3-319-30493-9_8.
- [45] J. Choi and Y. Tausczik, "Characteristics of collaboration in the emerging practice of open data analysis," *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Association for Computing Machinery, pp. 835–846, 2017. doi: 10.1145/2998181.2998265.
- [46] B. Palomo, L. Teruel, and E. Blanco-Castilla, "Data Journalism Projects Based on User-Generated Content. How La Nación Data Transforms Active Audience into Staff," *Digital Journalism*, vol. 7, no. 9. Routledge, pp. 1270–1288, 2019. doi: 10.1080/21670811.2019.1626257.
- [47] R. A. Handler and R. Ferrer Conill, "Open Data, Crowdsourcing and Game Mechanics. A case study on civic participation in the digital age," *Computer Supported Cooperative Work: CSCW: An International Journal*, vol. 25, no. 2–3. Springer Netherlands, pp. 153–166, 2016. doi: 10.1007/s10606-016-9250-0.
- [48] M. Kassen, "Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations," *Aslib Journal of Information Management*, vol. 70, no. 5. Emerald Group Holdings Ltd., pp. 518–537, 2018. doi: 10.1108/AJIM-11-2017-0250.
- [49] J. L. Boyles, "Laboratories for news? Experimenting with journalism hackathons," *Journalism*, vol. 21, no. 9. SAGE Publications Ltd, pp. 1338–1354, 2020. doi: 10.1177/1464884917737213.
- [50] S. Baack, "Practically Engaged: The entanglements between data journalism and civic tech," *Digit. J. Journal.*, vol. 6, no. 6, pp. 673–692, 2018, doi: 10.1080/21670811.2017.1375382.
- [51] A. K. Krotoski, "Data-driven research: Open data opportunities for growing knowledge, and ethical issues that arise," *Insights: the UKSG Journal*, vol. 25, no. 1. United Kingdom Serials Group, pp. 28–32, 2012. doi: 10.1629/2048-7754.25.1.28.
- [52] L. Camaj, J. Martin, and G. Lanosga, "The Impact of Public Transparency Infrastructure on Data Journalism: A Comparative Analysis between Information-Rich and Information-Poor Countries," *Digital Journalism*. Routledge, 2022. doi: 10.1080/21670811.2022.2077786.
- [53] F. Faini and M. Palmirani, "The right to know through the freedom of information and open data," *Proc. Eur. Conf. E-Gov. ECEG*, vol. 2016-January, pp. 54–62, 2016.
- [54] S. Baack, "Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism," *Big Data and Society*, vol. 2, no. 2. SAGE Publications Ltd, 2015. doi: 10.1177/2053951715594634.
- [55] L. Bernitz, M. R. Johannessen, K. N. Andersen, and J. Crusoe, "Parliamentary open data in scandinavia," *Computers*, vol. 8, no. 3. MDPI AG, 2019. doi: 10.3390/computers8030065.
- [56] B. T. Lawson, "Realizing the benefits of open government data: Journalists' coverage of the NHS winter crisis, 2016–17," *INFORMATION SOCIETY*, vol. 38, no. 1. TAYLOR & FRANCIS INC, 530 WALNUT STREET, STE 850, PHILADELPHIA, PA 19106 USA, pp. 25–35, Feb. 07, 2022. doi: 10.1080/01972243.2021.1998274.
- [57] J. A. Martin, L. Camaj, and G. Lanosga, "Scrape, Request, Collect, Repeat: How Data Journalists Around the World Transcend Obstacles to Public Data," *Journalism Practice*. Routledge, 2022. doi: 10.1080/17512786.2022.2142837.
- [58] F. Windhager, E. Mayr, G. Schreder, and M. Smuc, "Linked information visualization for linked open government data: A visual synthetics approach to governmental data and knowledge collections," *eJournal of eDemocracy and Open Government*, vol. 8, no. 2. Department for E-Governance and Administration, pp. 87–116, 2016. doi: 10.29379/jedem.v8i2.436.
- [59] A. Veglis and T. A. Maniou, "The mediated data model of communication flow: Big data and data journalism," *KOME*, vol. 6, no. 2. Hungarian Communication Studies Association, pp. 32–43, 2018. doi: 10.17646/KOME.2018.23.
- [60] P. Andrews and F. S. C. Da Silva, "Using parliamentary open data to improve participation," *ACM International Conference Proceeding Series*. Association for Computing Machinery, pp. 242–249, 2013. doi: 10.1145/2591888.2591933.
- [61] F. D. Davis F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–339.
- [62] M. Turner, B. Kitchenham, P. Brereton, S. Charters S., D. Budgen (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology*, 52(5), 463–479.
- [63] N. Marangunic, A., Granic, (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society* volume 14, 81–95.
- [64] E. Rogers. (2003). *Diffusion of Innovations – Fifth Edition*: The Free Press, New York, USA.
- [65] E. Loukis, D. Spinellis, and A. Katsigiannis. (2011). Barriers to the adoption of B2B e-marketplaces by large enterprises: lessons learnt from the Hellenic Aerospace Industry. *Information Systems Management*, 28(2), 130–146.
- [66] I. M. Al-Jabri and M. S. Sohail, (2012). Mobile Banking Adoption: Application of Diffusion of Innovation Theory. *Journal of Electronic Commerce Research*, 13(4), 373–385.
- [67] V. Venkatesh, J. Thong, and X. Xu (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1) 157–178.
- [68] G. Pozzi, F., Pigni, and C., Vitari, C. (2014). Affordance Theory in the IS Discipline: a Review and Synthesis of the Literature. In *Proceedings of Twentieth Americas Conference on Information Systems (AMCIS)*, Savannah, USA.
- [69] H., Wang, J., Wang, and Q., Tang (2018). A Review of Application of Affordance Theory in Information Systems. *Journal of Service Science and Management*, 11, 56–70.
- [70] J., Fromm, M., Mirbabaie, S., Stieglitz (2020). A Systematic Review of Empirical Affordance Studies: Recommendations for Affordance Research in Information Systems. In *Proceedings of the Twenty-Eighth European Conference on Information Systems (ECIS)*, Marrakesh, Morocco.
- [71] W., DeLone, and E., McLean (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1), pp. 60–95.
- [72] W., DeLone and E., McLean (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19(4), 9–30.
- [73] W., DeLone, and E., McLean, E. (2016). Information Systems Success Measurement. *Foundations and Trends in Information Systems*, 2(1), 1–116.