



Interacting with Data to Create Journalistic Stories: A Systematic Review

Daniele R. de Souza^{1(✉)}, Lorenzo P. Leuck², Caroline Q. Santos^{1,3}, Milene S. Silveira¹, Isabel H. Manssour¹, and Roberto Tietzmann²

¹ School of Technology, PUCRS, Porto Alegre, Brazil
danieleramossdesouza@gmail.com

² School of Communications, Arts and Design, PUCRS, Porto Alegre, Brazil

³ Department of Computer Science, UFVJM, Diamantina, Brazil

Abstract. With the increasing amount of data available in digital media, new professional practices emerged in journalism to gather, analyze and compute quantitative data that aims to yield potential pieces of information relevant to news reporting. The constant evolution of the field motivated us to perform a systematic review of the literature on data-driven journalism to investigate the state-of-art of the field, concerning the process, expressed by the “inverted pyramid of data journalism”. We aim to understand what are the techniques and tools that are currently being used to collect, clean, analyze, and visualize data. Also, we want to know what are the data sources that are presently being used in data journalism projects. We searched databases that include publications from both fields of computing and communication, and the results are presented and discussed through data visualizations. We identified the years with the highest number of publications, the publications’ authors and the fields of study. Then, we classified these works according to the changes in quantitative practices in journalism, and to the contributions in different categories. Finally, we address the challenges and potential research topics in the data journalism field. We believe the information gathered can be helpful to researchers, developers, and designers that are interested in data journalism.

Keywords: Data journalism · Computational journalism
Computer-assisted reporting · Systematic review

1 Introduction

With the increasing amount of data generated by different sources, and made available online, the journalism industry has sought changes in search of relevance. The demands for information by the online audiences have been continuously redefined as communication and information technologies have evolved, and this has given rise to a new term in this field: digital journalism. According to Kawamoto [7], the definition of digital journalism is changing along with the

change of technologies and the new ways of the area. The author conceptualizes this term as the “use of digital technologies to research, produce, and deliver news and information to an increasingly computer-literate audience”.

Data journalism, on the other hand, differs from the traditional journalism possibly “by the new possibilities that open up when it combines the nose for news with the ability to tell a compelling story, with the sheer scale and range of digital information now available” [6]. And those possibilities can come at any stage of the journalists work.

Beyond the power of technologies to support data journalism, researchers consider them has also influenced the news producing and the news-consuming process. Thus, digital journalism can also be seen as a combination of computing and computational thinking applied to the news production activities: data gathering, organization, sense-making and data dissemination [5]. Therefore, nowadays, journalists are faced with the need to acquire technological skills and learn how to use tools as Google Sheets, MS Excel, Open Refine, Tableau, among others.

In this study, we take a closer look at how researchers have discussed the relationship between access, manipulation, and presentation of large-scale data and journalistic stories. We want to understand what characterizes the data-driven journalism process and what elements or factors should be considered in this field. The main question this study attempts to answer is “*how are media professionals interacting with data to create journalistic stories and what is the current state of art of research in this field?*”. More specifically, our main contributions include:

- main techniques/tools that are being used to collect, clean, analyze, and visualize data;
- primary data sources that are being used in data journalism projects and research; and
- gaps identified in this field.

The remainder of this paper is organized as follows. In Sect. 2 we present the background and related work. The proposed methodology used for the systematic literature review is described in Sect. 3. The obtained results are presented in Sect. 4, and our final considerations and suggestions for future work are presented in Sect. 6.

2 Background

With the increasing amount of personal and public information available in digital spaces and networks, new professional practices emerged in journalism during the last decades to gather, analyze and compute quantitative data that aims to yield relevant information to reporting [3]. Journalism and computer science combined efforts as programmers approached newsrooms and journalists started acquiring programming skills. The constant evolution of the field

motivated us to perform a systematic review of data journalism, following the guidelines proposed by Kitchenham and Charters [9].

The quantitative practices of journalism can be defined as Computer-Assisted Reporting (CAR), Data Journalism (DJ) and Computational Journalism (CJ) [3]. CAR has its roots in Philip Meyers precision journalism, which was modeled using empirical methods for data gathering and statistical analysis to answer questions posed by reporting. It introduced the computational thinking to newsrooms and was considered an innovative form of investigative journalism up to the early 2000s. It was superseded by DJ as it goes beyond the idea of investigative reporting, focusing on data analysis, its presentation and the production of data-driven stories. CJ is also a descendant of CAR and differentiates itself from DJ because it is built around abstraction and automation, producing computable models, algorithms that can prioritize, classify, and filter information.

The term “inverted pyramid often defines news writing”, a writing architecture that proposes the presentation of the most relevant information at the beginning of the text, followed by hierarchically decreasing contents regarding interest [1]. This architecture is useful for news outlets because readers can quit reading at any time and still get the most important parts of the story [10]. The “inverted pyramid” become even more critical on the web and other digital media, since users spend 80% of their time looking at information above the page fold and, although users do sometimes scroll, they allocate only 20% of their attention to elements below the fold [11].

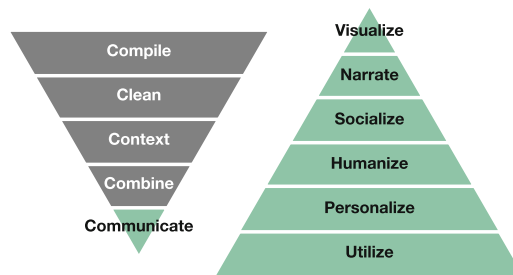


Fig. 1. The inverted pyramid of data journalism [1].

Conversely, Bradshaw [1] proposed the “inverted pyramid of data journalism” (Fig. 1) to explain the data journalism process to support those working with such content as journalists, developers, or designers. He presented as an inverted pyramid because it begins with a large amount of data that becomes increasingly focused to the point of communicating the results. It is composed of five stages: compile (gathering of data sources), clean (data preparation and error cleanup), context (inquire the sources, its biases, and purposes), combine (link data reporting with news story writing) and communicate (visualize, narrate, socialize, humanize, personalize and use) the results.

3 Methodology

According to Kitchenham [8], a systematic literature review (SLR) is a method for evaluation and interpretation of topics that are relevant to a research question, subject or event of interest. To conduct this study, we followed the guidelines of Kitchenham and Charters [9] to structure and organize our research. We defined our research goal and questions, and we established our research protocol. The main goal of this study is to investigate the state of art of data journalism research regarding its process as stated in the “inverted pyramid of data journalism”. Therefore, we designed the research questions as follows:

- **RQ1:** What are the techniques/tools that are used to collect data?
- **RQ2:** What are the techniques/tools that are used to clean data?
- **RQ3:** What are the techniques/tools that are used to analyze data?
- **RQ4:** What are the techniques/tools that are used to visualize data?
- **RQ5:** What are the data sources that are used in data journalism projects?

3.1 Search Strategy

This systematic review is focused on data journalism research conducted by researchers from computing or communications field of study or even by researchers from both areas working together. For this reason, we searched¹ databases that include publications from both fields of study (ACM Digital Library, IEEE Xplore, Elsevier ScienceDirect, and Scopus). Scopus contains Google Scholars top 10 academic journals in the communication area. Our search string was adapted according to the database, always including the four keywords we selected, which contemplate the three quantitative professional practices of journalism [3]: “Computer-Assisted Reporting”, “Data Journalism”, “Data-Driven Journalism”, and “Computational Journalism”. It is possible to see in Table 1 the list of consulted databases and their respective search strings.

3.2 Selection Strategy

Initially, we analyzed each publication retrieved from the initial search to remove duplicates and to include or exclude studies according to document type and language. The inclusion criteria were applied in the first filter: *(i)* English only; and *(ii)* conference and journal papers.

Publications with at least one of the following exclusion criteria were removed: duplicated, other languages, abstract only, book, and magazine. Subsequently, we conducted a **title, keywords, and abstract review**. In this phase, our inclusion criteria were:

- Fits into one of the three types of quantitative journalism (computer-assisted reporting, data journalism, or computational journalism) [3];

¹ Last search update: February 2nd 2018.

Table 1. Search string according to database

Database	Search string
ACM Digital Library	<i>(“computer-assisted reporting” OR “data journalism” OR “data-driven journalism” OR “computational journalism”) - Refined by: Journal and Proceedings.</i>
IEEE Xplore	<i>((((“computer-assisted reporting”) OR “data journalism”) OR “data-driven journalism”) OR “computational journalism”) - Refined by: Conferences Journals and Magazines.</i>
Science Direct	<i>TITLE-ABSTR-KEY(“computer-assisted reporting” OR “data journalism” OR “data-driven journalism” OR “computational journalism”)[All Sources(Computer Science,Social Sciences)]</i>
Scopus	<i>TITLE-ABS-KEY(“computer-assisted reporting” OR “data journalism” OR “data-driven journalism” OR “computational journalism”) AND (LIMIT-TO(DOCTYPE, “ar”) OR LIMIT-TO(DOCTYPE, “cp”)) AND (LIMIT-TO(SUBJAREA, “SOCI”) OR LIMIT-TO(SUBJAREA, “COMP”)) AND (LIMIT-TO(LANGUAGE, “English”)) AND (LIMIT-TO(SRCTYPE, “j”) OR LIMIT-TO(SRCTYPE, “p”))</i>

- Contributes to research on data journalism in the communications or computer science field of study.

Studies that did not fulfill one of these criteria were removed. Finally, we conducted a **full-text review**. In this phase, we performed a more detailed analysis of the papers’ content.

3.3 Data Extraction Strategy

We extracted for each study retrieved from the initial search the following data: year, title, keywords, abstract, authors, authors’ country, authors’ affiliation, publication name, and source database. In the full-text review, we extracted for each paper the following data, to answer our research questions and to classify all studies:

- Data collection tool/technique;
- Data cleaning tool/technique;
- Data analysis tool/technique;
- Data visualization tool/technique;

- Data source;
- Paper category;
- Type of quantitative journalism.

We analyzed the selected studies focusing on answering our research questions according to the data journalism process (data collection, cleaning, analysis, and visualization), as well as the data sources used in journalism projects. The collected dataset allowed us to identify which years had the highest number of publications and classified these works according to the changes in quantitative practices in journalism. We determined the publications' authors and their respective affiliations and countries of origin, as well as data concerning the publications' fields of study. We also did classify publication contributions in different categories, such as data journalism concepts, tools for journalists, application of data analysis/visualization techniques, and case studies. The results are then presented and discussed through visualizations.

4 Results

We applied the strategies of search, selection, and data extraction described in Sect. 3. Figure 2 presents the number of remaining papers according to each phase of the process. We obtained 273 papers from the initial search in the selected databases. First, we removed duplicates and applied a filter according to language and document type, resting 230 papers. After that, we executed a title keywords and abstract review, which left us with 111 papers that fulfilled our inclusion criteria. Finally, we performed a full-text review, remaining with 101 papers (Appendix A).

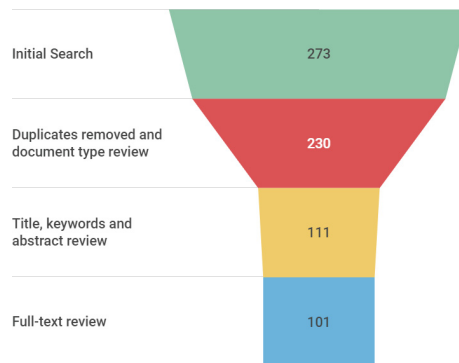


Fig. 2. Number of papers in each phase of the systematic review process.

Initially, we classified papers into six categories: data journalism concepts, case studies, new techniques, tools for journalists, application of existing techniques, and data journalism education. The treemap presented in Fig. 3 shows the final number of papers for each category.



Fig. 3. Selected papers classified by category.

Most of the papers discussed data journalism concepts, showing that this field is a new subject, still under development. In several cases, the research work employed in those studies serve as a link between communication professionals and academia, or as a way to understand significant cultural shifts and disruptive technologies. For this purpose, the primary method usually is literature reviews, as we can see in S35, S40, and S85.

The definition of the “case studies” category is very much like the one of the “data journalism concepts” category. However, in the former, the papers are related somewhat more to the data journalism practice than the theory itself. A closer look is held in the newsrooms routines, journalists workloads, communication demands, etc. S97, S99, and S101 show us a regular practice of this kind of study, the interview.

The “data journalism education” category is almost a meta-category, since it concerns not only the didactic and courses of data journalism, but also the way it is announced and understood. New modes of knowledge production in the area and experiences of postgraduate classes can be found in S62 and S80.

From the categories “tool for journalists”, “new techniques” and “application of existing techniques”, we could classify a series of systems, scripts, and interfaces used in data journalism. The difference among them is small, although utterly significant to our systematic review.

In the “tools for journalists” category, the authors present prototypes or final versions of systems developed by them to support specific stages or even the entire data journalism process [S37, S40]. On the other hand, the “new techniques” category encompasses scripts, methods or improvements to support any of the data journalism stages, as well as the entire data process [S22, S38]. The “application of existing techniques” category refers to the use of existing programs, not necessarily created by the authors. This is the most interdisciplinary

category since it is at the threshold of communication with information technology [S88, S90].

The following Subsects. 4.1 and 4.2 describe general information about the selected papers, and the answers of the research questions presented in the Sect. 3. Our primary research question will be answered in Sect. 5.

4.1 General Information About the Papers

In this subsection, we summarize some general data about the selected papers. Figure 4 shows the paper’s distribution according to publication year. As we can see, scientific production regarding data journalism began to grow significantly in recent years. Although the first papers were published in 1996, we can observe a linear growth from 2011 to 2014. From 2015, the number of published papers has grown exponentially. Only in the last three years, about 60% of the selected papers have been published. It is important to mention that, by the time we performed and updated the initial search, it is possible that not all the papers published in 2017 have been uploaded to the databases.

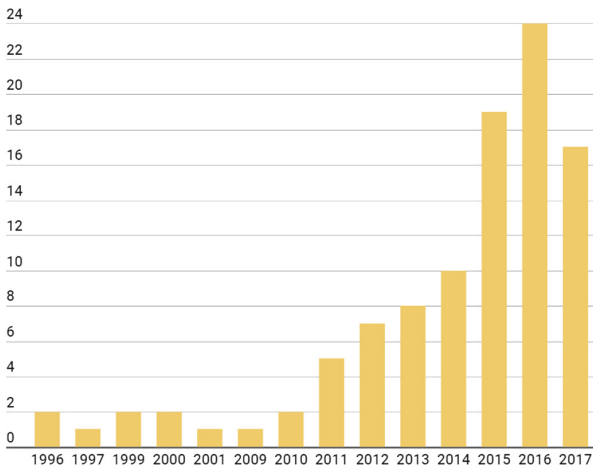


Fig. 4. Papers publication by year.

We found 195 distinct authors among the selected papers we analyzed, and 24 of them authored more than one article. From these, 7 published three or more papers, which we consider the top authors. Table 2 presents the list of authors who have more publications and the corresponding reference for each paper.

Analyzing only the selected papers authored by the top authors, we identified four networks of authors who recurrently publish together, which are shown in Fig. 5. A total of 18 papers generated the four networks, which comprise 35 authors with 89 connections between them (network density = 0,138, where 1

Table 2. Top authors of the selected papers.

Author	Papers	References
Diakopoulos N.	10	S9, S10, S11, S20, S42, S48, S58, S75, S81, S87
Li C.	8	S13, S22, S35, S37, S38, S52, S89, S100
Yang J.	6	S13, S22, S35, S37, S38, S89
Yu C.	5	S13, S22, S35, S38, S89
Wu Y.	4	S22, S35, S37, S89
Hassan N.	4	S35, S38, S52, S100
Agarwal P. K.	3	S22, S37, S89
Broussard M.	3	S43, S56, S62
Lewis S. C.	3	S28, S34, S45

represents a fully connected network). The thicker an edge of the graph, the more papers the connected authors published together.

The graphs showed in Fig. 5 represents how the top authors collaborated in their papers. As we can see, Diakopoulos has the most significant number of papers among our selection. However, his network has 22 connections and 14 co-authors (none of them in the list of top authors). Li C’s, on the other hand, has 15 co-authors and 57 connections, making it the biggest and most connected network in the graph. Lewis SC’s and Lewis J’s are equal in size, but Lewis J’s has three connections, one more than Lewis SC’s.

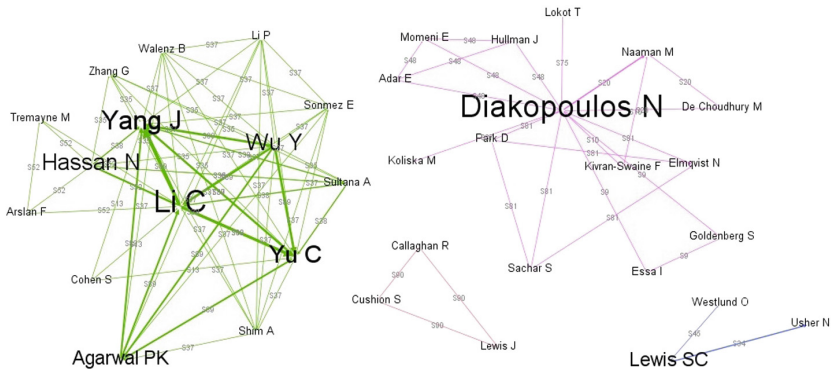


Fig. 5. Authors network.

Figure 6 shows an alternative visualization for the graph containing the authors network. In this case, we present the connections between authors from different fields of study by using different colors. The red edges starts from the

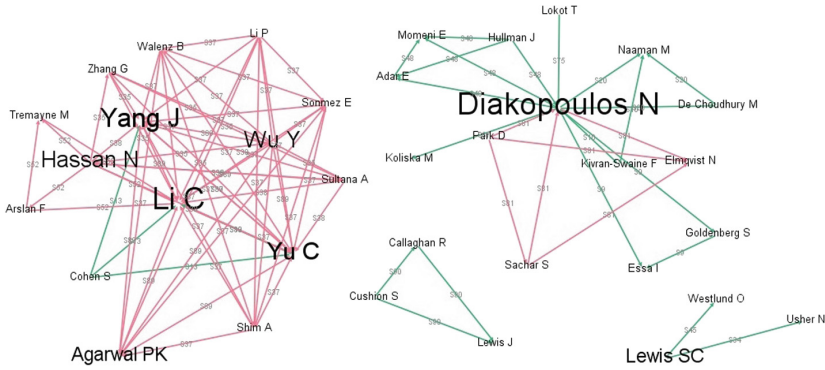


Fig. 6. Authors network according to field of study. (Color figure online)

nodes of the graph referring to the authors from the computing field of study and the green ones relate to the authors from the communications area. As we can see, the largest authors network comprises mainly papers published by authors from the computing area, except by Cohen S. The authors network generated by connections of Diakopoulos N. is mostly from the communications field of study.

We also analyzed the selected papers according to the country of authors' institutions (Fig. 7). There are 51 papers published by authors from the United States (49.03%), followed by 9 from the United Kingdom (8.65%) and 5 from Austria (4.8%). Despite the low percentage of publications from countries other than United States, papers from our sample indicate that the diversity in data journalism's research and practice is increasing.

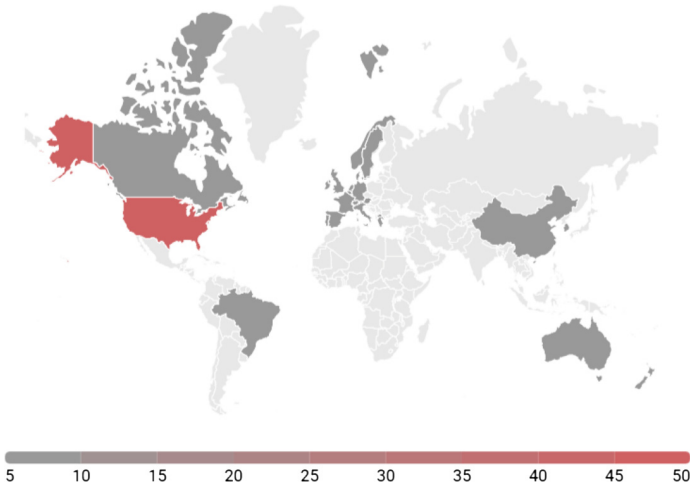


Fig. 7. Selected papers by country.

4.2 Answers to Research Questions

In this subsection, we describe the results of our SLR that answer the five research questions previous presented. The answers to the research questions were obtained by extracting and combining data from the 101 selected papers.

Several papers from our selection proposed new tools, both under development or ready to use, for supporting all the stages of the data journalism process. Some examples are Vox Civitas, SRSR, FactMinder, FactWatcher, Icheck, TATOOINE, News Context Project, TweetTalk, Readdit, The Openprocurement.mk, and YDS.

Data Collection

We identified 22 papers that helped us to answer our RQ1 (“*What are the techniques/tools that are used to collect data?*”). In S2, authors investigated the tools that daily newspapers were using in computer-assisted reporting from the application of an online survey questionnaire between December 1993 and March 1994. They mentioned that computers were becoming more and more valuable to reporters, not only for news writing but also for news gathering. For the authors, graphical user interfaces products, such as Windows and OS/2, that goes beyond DOS software, arrived to facilitate this task.

Regarding the scientific production of recent years, there are some papers presenting tools to support data collection (S17, S19, S20, S71, S76). Some of them were explicitly proposed for helping journalists in their practice (S20, S71). There are also some papers that describe new techniques or tools for the extracting information from different data sources (S10, S13, S26, S35, S37, S54, S59, S88, S91, S93). Some of them have the specific goal of supporting journalists in the discovery of news events (S10, S35, S91) or checking of facts (S26, S37). We also found one study proposing a database integration tool (S54).

We found only few papers in which authors report using existing tools and APIs for data collection. In S30 and S40, Twitter API was accessed directly or indirectly. In the latter case, a command line scraping tool was used to communicate with de API. S31, S41, and S58 mention the use of APIs from specific news sites, such as The Guardian Open Platform and The Times Community API. In S50, data was collected directly from open data government sites.

Data Cleaning

Our RQ2 (“*What are the techniques/tools that are used to clean data?*”) can be answered by 24 selected papers. Data cleaning methods and algorithms in S10, S20, S19, S26, S35, S37, S43, S54, S59, S71, S91, and S93 are used as part of closed systems, that is, tools developed for specific purposes that do not foresee changes or appropriation of functions.

In other instances, data cleaning occurs through independent steps, generated from well-established platforms and frameworks. Among these, S12 and S50 bring us examples of tabular data manipulations with Excel and Google Refine. In S2, we could find that researchers were using tools like XyWrite, WordPerfect, and Word. This tools can be considered Excel and Google Refine predecessors. An R script can be found in S40 and relational database management cases in

S14 and S89 with PostgreSQL and SQL. There are also examples in which the authors took an algorithmic approach: in-depth accounts (S15), classification and clustering engine (S7); text pre-processing (S30); clustering framework (S31); SpeakerRecognitionAPI, the Custom Recognition Intelligent Service (CRIS) and the Speech API (S72).

Data Analysis

There are 33 papers in our selection that address our RQ3 (*“What are the techniques/tools that are used to analyze data?”*). From these, several existing tools and techniques are used to support data analysis in a general way (S14, S15, S17, S19, S54, S71, S79 S93). Some papers deal with the analysis of specific types of content, such as text analysis (S10, S81, S88), audio analysis (S12, S72), and video analysis (S9). We found some papers that take advantage of users collaboration to analyze data (S72). Other papers take advantage of algorithms to automate the data analysis process. Among these, different analysis techniques are being used, such as natural language processing (S52, S55, S88) and clustering (S30, S76). Some papers are focused on sentiment analysis (S28, S30, S52).

We identified that some papers have the main goal of automatically discovering newsworthy themes in databases (S22, S35, S43, S52, S59, S61, S91), such as interesting facts (S22) and significant events (S61). Some papers aim to analyze data in order check facts (S26, S37, S89, S98). We also found some papers focused on the analysis of data retrieved from social media (S20, S30). Finally, there are only few papers in our selection that reported to use existing tools, from the most classic ones (S2) to some more modern and robust (S12, S40, S50).

Data Visualization

To answer RQ4 (*“What are the techniques/tools that are used to visualize data?”*), we analyzed all papers that referred to use visualizations techniques and/or tools to present data. In the other cases, the authors did not provide neither the technique’s name nor the tool used, and it was not possible to identify only by reading these papers. Among 101 papers, 31 mentioned the usage of some visualization technique or tool, which allowed us to extract what most used techniques in the data-driven journalism research (S2, S9, S10, S11, S12, S14, S17, S19, S20, S26, S30, S31, S35, S37, S40, S41, S43, S46, S49, S50, S54, S55, S59, S66, S71, S72, S76, S77, S81, S89, and S93). Most quoted visualization techniques are tables, graphs, charts, maps, and the tools quoted more than once time are D3.js and Tableau.

The papers ranged from theoretical discourse to the development and use of tools, besides discussing visualization techniques more used by them. It is interesting to observe that, in the midst of so many free and online tools that are available to support data journalists, we perceived that they are not widely used.

Data Sources

Our answer to RQ3 (“*What are the data sources that are used in data journalism projects?*”) is based on the sources referred in 43 papers (S9, S10, S11, S14, S15, S19, S20, S21, S22, S23, S26, S28, S30, S31, S35, S37, S38, S40, S41, S43, S46, S48, S49, S50, S52, S54, S55, S59, S61, S66, S71, S72, S76, S77, S79, S81, S88, S89, S90, S91, S93, S98, and S100). The others 58 did not mention anything about data sources.

Most of these papers reported different data sources used in data-driven journalism, ranging from media outlets, as BBC and The Guardian, to social media, like Twitter, Youtube, and Facebook. Besides that, government sources, political datasets, Wikipedia, NBA dataset, among others were also mentioned.

5 Discussion

Our goal with this study was to plot a state of art landscape on data-driven journalism research. In this section, we analyze the primary results that we obtained by conducting the systematic review in comparison to the main theories that support our work: the inverted pyramid of data journalism and its process’ stages [1], as well as the types of quantitative journalism [3] that evolved across time. We also discuss the implications for research in this field. Finally, we address the challenges and potential research topics in the data-driven journalism field.

This study seeks to contribute to the domain of digital journalism, or data-driven journalism, which has increased due to the advance of information and communications technologies. Although we were able to find some contributions on this field, we perceived this field as new and so there are still ongoing discussions on it.

Since the beginning of computers’ use in social and human sciences research/work, professionals and researchers have been benefited from facilities provided by technologies. The same happened to journalism. It started with computer-assisted reporting (CAR), in which technology facilitated the news producing and its workflow [12]. CAR was used as support to create investigative reporting, involving data collection, analysis, presentation, and archive [4]. Figure 8 shows the 101 papers (discussed in this study) by type of quantitative journalism across time.

The data we collected to answer the systematic review’s research questions converge with the chronological classification proposed by Coddington [3] for the different types of quantitative journalism. As we can see in the graphic presented in Fig. 8, Computer-assisted reporting was the forerunner of data journalism, using basic computing and statistical methods as an extension of reporters’ skills. In this period, graphic interfaces and computer programs such as Word and Excel helped journalism professionals from news gathering to news reporting. There are only a few papers about this type of quantitative journalism that were published from 1996 to 2001.

Data journalism came up with the possibility of analyzing large amounts of data in a way that was not possible since no human being would be able to

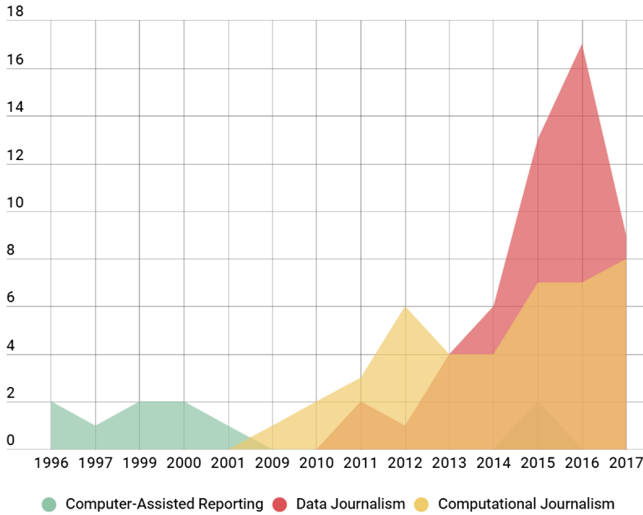


Fig. 8. Papers by type of quantitative journalism across time.

perform such task without machine help. Besides that, data journalism is characterized by its focus on data analysis and presentation, providing the readers with a new experience of interaction with the news stories, as well as with the opportunity for the public to collaborate with journalistic investigations in the data gathering stage (a process named as crowdsourcing). The rise of such activity poised discussions on data transparency from public and private players and to have access to government data portals. Scientific production regarding data journalism began to grow after 2007 and remains on the rise.

Computational journalism, in turn, arose from the direct influence of computer science and concerns the use of algorithms for the automation of data journalism processes. It is related to discussions about its implementation in the newsrooms and the resistance of some journalists who feel threatened, primarily by the news bots. As Fig. 8 shows, papers regarding computational journalism are growing exponentially in recent years. We believe there still plenty of room for conducting studies in this area considering the benefits they can bring to the practice of journalism today.

In the journalism field, graphics have been used for a long time to present statistical and non-statistical data and to allow audiences to view the information guided by the author [2], a visual solution called infographic. Until not long ago, some researchers compared infographics with data visualization, claiming that this latest emphasizes the interaction, which allows audiences to conduct a customized analysis of the data while the first could be just an editorial decision. However, there is a trend of journalists and researchers, as Cairo [2], to understand infographics and data visualizations as an organic continuum field of study, and, nowadays, there is little difference between them.

In this systematic review, we perceived that there is not much research that focuses on discussing visualization techniques or ideal tools for journalism, and not even research offering guidance to applying visualization in data-driven journalism. Thus, we believe that there is a gap in this field about the ways media outlets could incorporate visualization in the digital news. Likewise, media professionals, journalists, and researchers should improve their studies in a documentation and description of techniques and tools used by them, enabling future professionals to continue their work.

To answer our primary research question (*“how are media professionals interacting with data to create journalistic stories and what is the current state of art of research in this field?”*), we found that the journalistic skills to deal with large volumes of data are improving over time. Also, nowadays, they have a range of tools that support their work activities, as we identified in this SLR.

Considering these changes in the journalist’s profession, there is a need to redefine the curricula of journalism courses, including mathematics and programming disciplines on it to better prepare the new generation of journalism professionals for the challenges of the data age. It is also interesting to unite communications and computing areas even further, creating repositories with tools to support the data journalism process in its different stages (compile, clean, context, combine, and communicate) [1]. Cohen et al. [S12] describe this idea:

“Journalists need to partner with computer scientists, application developers, and hardware engineers. For decades, the computing community has empowered individuals to seek information, improving their lives in the process. Few fields have done more to give citizens the tools they need to govern themselves. Few fields today need computer scientists more than public interest journalism.”

6 Final Considerations

This study presented a systematic review of the state-of-the-art research on data-driven journalism. The selected papers encompass studies conducted by researchers from computer science and communication fields or even by researchers from both areas working together. From the 273 papers retrieved from the automated search, we selected 101 that we considered relevant for investigation. These papers address several topics concerning data-driven journalism.

We analyzed the collected data in order to answer our primary research question (*“how are media professionals interacting with data to create journalistic stories and what is the current state of art of research in this field?”*). Besides that, we answered some systematic review’s research questions according to the process’ stages (data collection, cleaning, analysis, and visualization) and sources used in data journalism projects.

Our findings showed that the relationship between journalists and data to create news stories is still seen as a recent practice, albeit a growing one. Media

outlets increasingly apply the data collection, cleaning, analysis, and visualization process as an efficient way to tell a persuasive news story. Therefore, computational tools have been incorporated into the news producing routine.

Our contributions were: *(i)* presentation of an overview of existing research in data-driven journalism; *(ii)* identification of main techniques/tools that are being used to collect, clean, analyze, and visualize data in this field; *(iii)* the primary data sources that are being used in data journalism projects and research; and *(iv)* the possible gaps in this field.

We believe the information gathered in this systematic review can be helpful to researchers, developers, and designers that are interested in data journalism, considering that different types of users may benefit from such visualizations, either as journalists or the general public.

Regarding future works in the field as mentioned earlier, we envision this text to be capable of becoming a methodological blueprint for updates in coming years. As the area of data-driven journalism continues to evolve, aligned with the academic and professional research on the field, we foresee the need to update the literature review in a few years time. Other possible implementations, as the use of our methodology to focus specifically on texts from a country or a timeframe, suggest future fields of development. Also, since the academic datasets were already collected, the creation of interactive and responsive visualizations could help scholars and students to perform a visual analysis and search of the works.

Acknowledgements. This work was partially supported by PUCRS (Edital N 05/2017 - Programa de Apoio a Integração entre Áreas/PRAIAS). The authors also acknowledge the support of CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) in this work.

Appendix A - List of selected studies

- S1** - Parsons P., Johnson R.B.: ProfNet: A Computer-Assisted Reporting Bridge to Academia. In: Newspaper Research Journal, Vol 17, Issue 3-4, pp. 29–38 (1996)
- S2** - Garisson B.: Tools Daily Newspapers Use in Computer-Assisted Reporting. In: Newspaper Research Journal, Vol 17, Issue 1-2, pp. 1131–126 (1996)
- S3** - Williams P., Nicholas D.: Journalists, news librarians and the Internet. In: New Library World, Vol 98, Issue 6, pp. 217–223 (1997)
- S4** - Garrison B.: Newspaper size as factor in use of computers for newsgathering. In: Newspaper Research Journal, Vol 20, Issue 3, pp. 72–185 (1999)
- S5** - Martin H.: The changing information environment in the media: Case study the Guardian/Observer. In: Aslib Proceedings, Vol 51, Issue 3, pp. 91–196 (1999)
- S6** - Mayo J., Leshner G.: Assessing the credibility of computer-assisted reporting. In: Newspaper Research Journal, Vol 21, Issue 4, pp. 68–182 (2000)
- S7** - Maier S.R.: Digital diffusion in newsrooms: The uneven advance of computer-assisted reporting. In: Newspaper Research Journal Vol 21, Issue 2, pp. 95–110 (2000)
- S8** - Deuze M.: Online journalism: Modelling the first generation of news media on the World Wide Web. In: First Monday, Vol 6, Issue 10 (2001)
- S9** - Diakopoulos N., Goldenberg S., Essa I.: Videolyzer: Quality Analysis of Online Informational Video for Bloggers and Journalists. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09), pp. 799–808 (2009)
- S10** - Diakopoulos N., Naaman M., Kivran-Swaine F.: Diamonds in the rough: Social media visual analytics for journalistic inquiry. In: IEEE Symposium on Visual Analytics Science and Technology, pp. 115–122 (2010)
- S11** - Diakopoulos N.: Game-y information graphics. In: Extended Abstracts on Human Factors in Computing Systems (CHI EA'10), pp. 3595–3600 (2010)
- S12** - Cohen S., Hamilton J. T., Turner F.: Computational journalism. In: Communications of the

ACM, Vol 54, Issue 10, pp. 66–71 (2011)

S13 - Cohen S., Li C., Yang J., Yu C.: Computational journalism: A call to arms to database researchers. In: 5th Biennial Conference on Innovative Data Systems Research, Conference Proceedings (CIDR 2011), pp. 148–151 (2011)

S14 - Pulimood S. M., Shaw D., Lounsberry E.: Gumshoe: A model for undergraduate computational journalism education. In: Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE'11), pp. 529–534 (2011)

S15 - Wagner E. J., Lin J.: In-depth Accounts and Passing Mentions in the News: Connecting Readers to the Context of a News Event. In: Proceedings of the 2011 iConference, pp. 790–791 (2011)

S16 - Hollings J.: The informed commitment model: Best practice for journalists engaging with reluctant, vulnerable sources and whistle-blowers. In: Pacific Journalism Review: Te Koaboa, Vol 17, Issue 1, pp. 67–89 (2011)

S17 - Francisco-Revilla L., Figueira A.: Adaptive Spatial Hypermedia in Computational Journalism. In: Proceedings of the 23rd ACM conference on Hypertext and social media (HT'12), pp. 313–314 (2012)

S18 - Weber W., Rall H.: Data visualization in online journalism and its implications for the production process. In: 16th International Conference on Information Visualisation, pp. 349–356 (2012)

S19 - Francisco-Revilla L.: Digital Libraries for Computational Journalism. In: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries (JCDL'12), pp. 365–366 (2012)

S20 - Diakopoulos N., De Choudhury M., Naaman M.: Finding and assessing social media information sources in the context of journalism. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12), pp. 2451–2460 (2012)

S21 - Pellegrini, T.: Integrating linked data into the content value chain: a review of news-related standards, methodologies and licensing requirements. In: Proceedings of the 8th International Conference on Semantic Systems, pp. 94–102 (2012)

S22 - Wu Y., Agarwal P.K., Li C., Yang J., Yu C.: On "one of the few" objects. In: 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12), pp. 1487–1495 (2012)

S23 - Pellegrini, T.: Semantic Metadata in the News Production Process: Achievements and Challenges. In: Proceeding of the 16th International Academic MindTrek Conference (MindTrek'12), pp. 125–133 (2012)

S24 - Lesage F., Hackett R. A.: Between objectivity and openness: the mediality of data for journalism. In: Media and Communication, Vol 2, Issue 2, pp. 42–54 (2013)

S25 - Parasio S., Dagiral E.: Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. In: New Media Society, Vol 15, Issue 6, pp. 853–871 (2013)

S26 - Goasdou F., Leblay J., Karanasos K., Manolescu I., Katsis Y., Zampetakis S.: Fact checking and analyzing the Web. In: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD'13), pp. 997–1000 (2013)

S27 - Lewis S.C., Usher N.: Open source and journalism: Toward new frameworks for imagining news innovation. In: Media, Culture Society, Vol 35, Issue 5, pp. 602–619 (2013)

S28 - Thelwall M., Buckley K.: Topic-based sentiment analysis for the social web: The role of mood and issue-related words. In: Journal of the Association for Information Science and Technology, Vol 64, Issue 8, pp. 1608–1617 (2013)

S29 - Anderson C. W.: Towards a sociology of computational and algorithmic journalism. In: New Media Society, Vol 15, Issue 7, pp. 1005–1021 (2013)

S30 - Lin Y.-R., Margolin D., Keegan B., Lazer D.: Voices of victory: A computational focus group framework for tracking opinion shift in real time. In: Proceedings of the International Conference on World Wide Web (WWW'13), pp. 737–748 (2013)

S31 - Gall M., Renders J.-M., Karstens E.: Who Broke the News?: An Analysis on First Reports of News Events. In: Proceedings of the 22nd International Conference on World Wide Web (WWW'13), pp. 855–862 (2013)

S32 - Lesage F., Hackett R.A.: Between objectivity and openness: The mediality of data for journalism. In: Media and Communication, Vol 2, Issue 2, pp. 42–54 (2014)

S33 - Griffin P.: Big news in a small country: developing independent public interest journalism in NZ. In: Pacific Journalism Review, Vol 20, Issue 1, pp. 11–34 (2014)

S34 - Lewis S.C., Usher N.: Code, Collaboration, And The Future Of Journalism: A case study of the Hacks/Hackers global network. In: Journalism and Mass Communication, Vol 2, Issue 3, pp. 383–393 (2014)

S35 - Hassan N., Sultana A., Wu Y., Zhang G., Li C., Yang J., Yu C.: Data in, fact out: Automated monitoring of facts by FactWatcher. In: 40th International Conference on Very Large Data Bases (VLDB 2014), pp. 155–1560 (2014)

S36 - Appelgren E., Nygren G.: Data Journalism in Sweden: Introducing new methods and genres of journalism into old organizations. In: Digital Journalism, Vol 2, Issue 3, pp. 394–405 (2014)

S37 - Wu Y., Walenz B., Li P., Shim A., Sonmez E., Agarwal P. K., Li C., Yang J. Yu C.: iCheck: Computationally Combating Lies, D–ned Lies, and Statistics. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD'14), pp. 1063–1066 (2014)

S38 - Sultana A., Hassan N., Li C., Yang J., Yu C.: Incremental discovery of prominent situational facts. In: 2014 IEEE 30th International Conference on Data Engineering (ICDE), pp. 112–123 (2014)

- S39** - Gynnild A.: Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. In: *Journalism*, Vol 15, Issue 6, pp. 713–730 (2014)
- S40** - Regattieri L.L., Rockwell G., Chartier R., Windsor J.: TweetViz: Following Twitter hashtags to support storytelling. In: *HT (Doctoral Consortium/Late-breaking Results/Workshops)*, Vol 1210 (2014)
- S41** - Corby T.: Visualizing the news: Mutant barcodes and geographies of conflict. In: *Leonardo*, Vol 47, Issue, pp. 84–85 (2014)
- S42** - Diakopoulos N.: Algorithmic Accountability: Journalistic investigation of computational power structures. In: *Digital Journalism*, Vol 3, Issue 3, pp. 398–415 (2015)
- S43** - Broussard M.: Artificial Intelligence for Investigative Reporting: Using an expert system to enhance journalists ability to discover original public affairs stories. In: *Digital Journalism*, Vol 3, Issue 6, pp. 814–831 (2015)
- S44** - Anderson C. W.: Between the Unique and the Pattern: Historical tensions in our understanding of quantitative journalism. In: *Digital Journalism*, Vol 3, Issue 3, pp. 349–363 (2015)
- S45** - Lewis S.C., Westlund O.: Big Data and Journalism: Epistemology, expertise, economics, and ethics. In: *Digital Journalism*, Vol 3, Issue 3, pp. 447–466 (2015)
- S46** - Graham C.: By the Numbers: Data Journalism Projects as a Means of Teaching Political Investigative Reporting. In: *Asia Pacific Media Educator*, Vol 25, Issue 2, pp. 247–261 (2015)
- S47** - Coddington M.: Clarifying Journalisms Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. In: *Digital Journalism*, Vol 3, Issue 3, pp. 331–348 (2015)
- S48** - Hullman J., Diakopoulos N., Momeni E., Adar E.: Content, context, and critique: Commenting on a data visualization Blog. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work Social Computing (CSCW'15)*, pp. 1170–1175 (2015)
- S49** - Knight M.: Data journalism in the UK: A preliminary analysis of form and content. In: *Journal of Media Practice*, Vol 16, Issue 1, pp. 55–72 (2015)
- S50** - Plaue C., Cook L.R.: Data journalism: Lessons learned while designing an interdisciplinary service course. In: *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE'15)*, pp. 126–131 (2015)
- S51** - Parasie S.: Data-Driven Revelation?: Epistemological tensions in investigative journalism in the age of big data. In: *Digital Journalism*, Vol 3, Issue 3, pp. 364–380 (2015)
- S52** - Hassan N., Li C., Tremayne M.: Detecting Check-worthy Factual Claims in Presidential Debates. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM'15)*, pp. 1835–1838 (2015)
- S53** - Young M. L., Hermida A.: From Mr. and Mrs. Outlier To Central Tendencies: Computational journalism and crime reporting at the Los Angeles Times. In: *Digital Journalism*, Vol 3, Issue 3, pp. 381–397 (2015)
- S54** - Bonaque R., Cao T.D., Cautis B., Goasdou F., Letelier J., Manolescu I., Mendoza O., Ribeiro S., Tannier X., Thomazo M.: Mixed-instance querying: A lightweight integration architecture for data journalism. In: *Proceedings of the VLDB Endowment*, Vol 9, Issue 13, pp. 1513–1516 (2015)
- S55** - Berendt B.: Power to the agents?! in the #WebWeWant, people will critically engage with data - And data journalism can help them want to do this. In: *Joint Proceedings of the 5th International Workshop on Using the Web in the Age of Data (USEWOD '15) and the 2nd International Workshop on Dataset PROFiling and Federated Search for Linked Data (PROFILES '15) co-located with the 12th European Semantic Web Conference (ESWC 2015)*, pp. 29–31 (2015)
- S56** - Broussard M.: Preserving news apps present huge challenges. In: *Newspaper Research Journal*, Vol 36, Issue 3, pp. 299–313 (2015)
- S57** - Rodriguez M.T., Nunes S., Devezas T.: Telling stories with data visualization. In: *Proceedings of the 2015 Workshop on Narrative Hypertext (NHT'15)*, pp. 7–11 (2015)
- S58** - Diakopoulos N.: The editor's eye: Curation and comment relevance on the New York Times. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work Social Computing (CSCW'15)*, pp. 1153–1157 (2015)
- S59** - Birnbaum L., Boon M., Bradley S., Wilson J.: The news context project. In: *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion (UIU Companion'15)*, pp. 5–8 (2015)
- S60** - De Maeyer J., Libert M., Domingo D., Heinderyckx F., Le Cam F.: Waiting for Data Journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. In: *Digital Journalism*, Vol 3, Issue 3, pp. 432–446 (2015)
- S61** - Kim S., Oh J., Lee J.: Automated news generation for TV program ratings. In: *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX'16)*, pp. 141–145 (2016)
- S62** - Broussard M.: Big Data in Practice: Enabling computational journalism through code-sharing and reproducible research methods. In: *Digital Journalism*, Vol 4, Issue 2, pp. 266–279 (2016)
- S63** - Griffin R.J., Dunwoody S.: Chair support, faculty entrepreneurship, and the teaching of statistical reasoning to journalism undergraduates in the United States. In: *Journalism*, Vol 17, Issue 1, pp. 97–118 (2016)
- S64** - Hu Y., Lin Y.-R., Luo J.: Collective sensemaking via social sensors: Extracting, profiling, analyzing, and predicting real-world events. In: *Proceedings of the 22nd ACM SIGKDD International*

- Conference on Knowledge Discovery and Data Mining (KDD'16), pp. 2127–2128 (2016)
- S65** - Davies K., Cullen T.: Data Journalism Classes in Australian Universities: Educators Describe Progress to Date. In: *Asia Pacific Media Educator*, Vol 26, Issue 2, pp. 132–147 (2016)
- S66** - Tabary C., Provost A.-M., Trottier A.: Data journalism's actors, practices and skills: A case study from Quebec. In: *Journalism*, Vol 17, Issue 1, pp. 66–84 (2016)
- S67** - Appelgren E.: Data Journalists Using Facebook: A Study of a Resource Group Created by Journalists, for Journalists. In: *Nordicom Review*, Vol 37, Issue 1, pp. 1–14 (2016)
- S68** - Zhu H., Lee Y. W., Rosenthal A. S.: Data Standards Challenges for Interoperable and Quality Data. In: *Journal of Data and Information Quality (JDIQ) - Challenge Papers, Regular Papers and Experience Paper*, Vol 7, Issue 1-2, Article No. 4 (2016)
- S69** - Felle T.: Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. In: *Journalism*, Vol 17, Issue 1, pp. 85–96 (2016)
- S70** - Splendore S., Di Salvo P., Eberwein T., Groenhard H., Kus M., Porlezza C.: Educational strategies in data journalism: A comparative study of six European countries. In: *Journalism*, Vol 17, Issue 1, pp. 138–152 (2016)
- S71** - Shehu V., Mijushkovic A., Besimi A.: Empowering Data Driven Journalism in Macedonia. In: *Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics*, Article No. 49 (2016)
- S72** - Sidiropoulos E.A., Konstantinidis E. I., Veglis A. A.: Framework of a collaborative audio analysis and visualization tool for data journalists. In: *11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, pp. 156–160 (2016)
- S73** - Zwinger S., Zeiller M.: Interactive infographics in German online newspapers. In: *Proceedings of the 9th Forum Media Technology*, pp. 54–64 (2016)
- S74** - Hewett J.: Learning to teach data journalism: Innovation, influence and constraints. In: *Journalism*, Vol 17, Issue 1, pp. 119–137 (2016)
- S75** - Lokot T., Diakopoulos N.: News Bots: Automating news and information dissemination on Twitter. In: *Digital Journalism*, Vol 4, Issue 6, pp. 682–699 (2016)
- S76** - Le Borgne Y.-A., Homolova A., Bontempi G.: OpenTED browser: Insights into European Public Spendings. In: *1st Workshop on Data Science for Social Good* (2016)
- S77** - Salovaara I.: Participatory Maps: Digital cartographies and the new ecology of journalism. In: *Digital Journalism*, Vol 4, Issue 7, pp. 827–837 (2016)
- S78** - Gertrudis-Casado M.-C., Grtrudix-Barrio M., lvarez-Garca S.: Professional information skills and open data. Challenges for citizen empowerment and social change. In: *Comunicar*, Vol 24, Issue 47, p. 39 (2016)
- S79** - Orellana-Rodriguez C., Greene D., Keane M. T.: Spreading the news: How can journalists gain more engagement for their tweets?. In: *Proceedings of the 8th ACM Conference on Web Science (WebSci'16)*, pp. 107–116 (2016)
- S80** - Yang F., Du Y. R.: Storytelling in the Age of Big Data: Hong Kong Students Readiness and Attitude towards Data Journalism. In: *Asia Pacific Media Educator*, Vol 26, Issue 2, pp. 148–162 (2016)
- S81** - Park D., Sachar S., Diakopoulos N., Elmqvist N.: Supporting comment moderators in identifying high quality online news comments. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*, pp. 1114–1125 (2016)
- S82** - La-Rosa L., Sandoval-Martín T.: The transparency laws insufficiency for data journalism practices in Spain. In: *Revista Latina de Comunicacin Social*, Vol 71, pp. 1208–1229 (2016)
- S83** - Heravi B.R., Harrower N.: Twitter journalism in Ireland: sourcing and trust in the age of social media. In: *Information, Communication Society*, Vol 19, Issue 9, pp. 1194–1213 (2016)
- S84** - Borges-Rey E.: Unravelling Data Journalism: A study of data journalism practice in British newsrooms. In: *Journalism Practice*, Vol 10, Issue 7, pp. 833–843 (2016)
- S85** - Bucher T.: Machines dont have instincts: Articulating the computational in journalism. In: *New Media Society*, Vol 19, Issue 6, pp. 918–933 (2017)
- S86** - S. Zwinger; J. Langer; M. Zeiller: Acceptance and Usability of Interactive Infographics in Online Newspapers. In: *21st International Conference Information Visualisation (IV)*, pp. 176–181 (2017)
- S87** - Diakopoulos N., Koliska M.: Algorithmic Transparency in the News Media. In: *Digital Journalism*, Vol 5, Issue 7, pp. 809–828 (2017)
- S88** - Boon M.: Augmenting media literacy with automatic characterization of news along pragmatic dimensions. In: *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17)*, pp. 49–52 (2017)
- S89** - Wu Y., Agarwal P.K., Li C., Yang J., Yu C.: Computational fact checking through query perturbations. In: *ACM Transactions on Database Systems (TODS) - Invited Paper from ICDT 2014, Invited Paper from EDBT 2015, Regular Papers and Technical Correspondence*, Vol 42, Issue 1, Article No. 4 (2017)
- S90** - Cushion S., Lewis J., Callaghan R.: Data Journalism, Impartiality And Statistical Claims: Towards more independent scrutiny in news reporting. In: *Journalism Practice*, Vol 11, Issue 10, pp. 1198–1215 (2017)
- S91** - Fan Q., Li Y., Zhang D., Tan K.-L.: Discovering newsworthy themes from sequenced data: A step towards computational journalism. In: *EEE Transactions on Knowledge and Data Engineering*,

Vol 29, Issue 7, pp. 1398–1411 (2017)

S92 - Langer J., Zeiller M.: Evaluation of the user experience of interactive infographics in online newspapers. In: 10th Forum Media Technology (2017)

S93 - Brolchin N., Porwol L., Ojo A., Wagner T., Lopez E.T., Karstens E.: Extending open data platforms with storytelling features. In: Proceedings of the 18th Annual International Conference on Digital Government Research (dg.o'17), pp. 48–53 (2017)

S94 - Wihbey J.: Journalists Use of Knowledge in an Online World: Examining reporting habits, sourcing practices and institutional norms. In: Journalism Practice, Vol 11, Issue 10, pp. 1267–1282 (2017)

S95 - Bounegru L., Venturini T., Gray J., Jacomy M.: Narrating Networks: Exploring the affordances of networks as storytelling devices in journalism. In: Digital Journalism, Vol 5, Issue 6, pp. 699–730 (2017)

S96 - Boyles J.L., Meyer E.: Newsrooms accommodate data-based news work. In: Newspaper Research Journal, Vol 38, Issue 4, pp. 428–438 (2017)

S97 - Tandoc E.C., Jr., Oh S.-K.: Small Departures, Big Continuities?: Norms, values, and routines in The Guardians big data journalism. In: Journalism Studies, Vol 18, Issue 8, pp. 997–1015 (2017)

S98 - Patwari A., Goldwasser D., Bagchi S.: TATHYA: A Multi-Classifer System for Detecting Check-Worthy Statements in Political Debates. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM'17), pp. 2259–2262 (2017)

S99 - Lopez-Garcia X., Rodriguez-Vzquez A.-I., Pereira-Faria X.: Technological skills and new professional profiles: Present challenges for journalism. In: Comunicar, Vol 25, Issue 53, pp. 81–90 (2017)

S100 - Hassan N., Arslan F., Li C., Tremayne M.: Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17), pp. 1803–1812 (2017)

S101 - Thurman N., Drr K., Kunert J.: When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences. In: Digital Journalism, Vol 5, Issue 10, pp. 1240–1259 (2017)

References

1. Bradshaw, P.: The inverted pyramid of data journalism. *Online Journalism Blog* 7 (2011)
2. Cairo, A.: *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders (2012)
3. Coddington, M.: Clarifying journalism's quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digit. Journalism* 3(3), 331–348 (2015)
4. Cox, M.: The development of computer-assisted reporting. Informe presentado en Association for Education in Journalism and Mass Communication. EEUU: Universidad de Carolina del Norte, Chapel Hill (2000)
5. Diakopoulos, N.: A functional roadmap for innovation in computational journalism. *Nick Diakopoulos* (2011)
6. Gray, J., Chambers, L., Bounegru, L.: *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. O'Reilly Media, Inc., Sebastopol (2012)
7. Kawamoto, K.: *Digital Journalism: Emerging Media and the Changing Horizons of Journalism*. Rowman & Littlefield Publishers, Lanham (2003)
8. Kitchenham, B.: Procedures for performing systematic reviews. *Keele University, Keele*, vol. 33, pp. 1–26 (2004)
9. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Technical report, Ver. 2.3 EBSE Technical report. EBSE (2007)
10. Nielsen, J.: *Inverted Pyramids in Cyberspace*. Nielsen Norman Group (1996)
11. Nielsen, J.: *Scrolling and Attention*. Nielsen Norman Group (2010)
12. Royal, C.: The journalist as programmer: a case study of the New York times interactive news technology department. In: *International Symposium on Online Journalism* (2010)