The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork

Julian Ausserhofer
University of Vienna, Austria; Alexander von Humboldt Institute for Internet and Society, Germany

Robert Gutounig
FH JOANNEUM University of Applied Sciences, Austria

Michael Oppermann
University of Vienna, Austria

Sarah Matiasek
FH JOANNEUM University of Applied Sciences, Austria

Eva Goldgruber
FH JOANNEUM University of Applied Sciences, Austria

Abstract
This article explores the existing research literature on data journalism. Over the past years, this emerging journalistic practice has attracted significant attention from researchers in different fields and produced an increasing number of publications across a variety of channels. To better understand its current state, we surveyed the published academic literature between 1996 and 2015 and selected a corpus of 40 scholarly works

Corresponding author:
Julian Ausserhofer, Department of Communication, University of Vienna, Währingerstraße 29, 1090 Vienna, Austria.
Email: julian.ausserhofer@hiig.de

that studied data journalism and related practices empirically. Analyzing this corpus with both quantitative and qualitative techniques allowed us to clarify the development of the literature, influential publications, and possible gaps in the research caused by the recurring use of particular theoretical frameworks and research designs. This article closes with proposals for future research in the field of data-intensive newswork.

Keywords

Citation analysis, computational journalism, computer-assisted reporting, data journalism, data-driven journalism, literature review, literature survey, meta-analysis, research synopsis, research synthesis

Introduction

In late August 2010, around 50 international journalists and researchers convened in Amsterdam for a new roundtable conference. The participants, who included leading personnel from organizations such as the *New York Times*, the *Guardian*, the *Financial Times*, the *Open Knowledge Foundation*, and *Hacks/Hackers*, presented their works and discussed the state of a craft which had a few months earlier been labeled 'data journalism' (Rogers, 2008, as cited in Knight, 2015: 55-56). The publication that documented this meeting – 'Data-driven journalism: What is there to learn?' (European Journalism Centre, 2010) – helped to shape the early discourse around a practice that has since been adopted by individuals and organizations all over the world. As academics across a variety of disciplines have taken notice, the research literature has grown tremendously – far beyond journalism studies itself. In this article, we revive the motto 'What is there to learn?' for a meta-analysis of the research publications on the topic.

Anderson (2013) suggested 'six approaches to a sociology of computational journalism' (p. 1010), outlining a framework to investigate a phenomenon that, until then, had been explored primarily by computer scientists and practitioners active in the field. His proposal seemed driven by a discontent with the 'internalist tendencies at [the] early stage of academic research' (p. 1007) that often defaulted toward hopeful or fearful statements of how the practice would affect the future of journalism. To counter this trend, Anderson recommended studies that used political, cultural, organizational, technological, or field perspectives for their research – in effect following and extending Schudson's (2005) typology of the sociology of news. While the extent of his influence is unclear, the field changed rapidly: only 2 years later, together with Fink, Anderson spoke of 'an explosion in data journalism-oriented scholarship' (Fink and Anderson, 2015: 476). Lewis (2015) also attested to a 'rapidly growing body' (p. 322) of scientific studies investigating the field (as mentioned in Loosen et al., 2015: 2).

What was all this research about? Empirically, 'the first wave of data journalism research' (Uskali and Appelgren, 2015) produced a number of studies, detailing the national scenes and data desks of Western newsrooms. Many covered the practice in one (or more) media organization(s) over several weeks and were descriptive, examining the organizational culture, newsroom structures, the epistemologies of data journalists, and

the characteristics of data-intensive pieces. Typical of the 'first wave' research, the findings were rather detailed and led to a fragmented publication landscape. Both then and now, the interdisciplinary nature of the field also seems to reinforce this fragmentation, for while it stands to benefit from broad academic engagement, the fact that researchers publish in outlets important to their discipline means that many efforts happened (and continue to happen) in parallel, resulting in similar findings.

Research questions

To address the granularity and fragmentation of data journalism research, this article adheres to the guidelines of structured literature reviews. It positions itself in the tradition of other meta-studies devoted to special areas of journalism (e.g. Neuberger et al., 2007, who surveyed the potential of blogs for journalism; or Steensen and Ahva, 2015, who collected the theoretical underpinnings of journalism research). In a related meta-analysis, Diakopoulos (2012) analyzed and explored 'the potential for technical innovation in journalism' (p. 2). By systematically analyzing the computer science literature that covered journalism, he found areas previously neglected by the computer science community that he suggested could further its development. This article is complementary to Diakopoulos' publication. While he explored the potential for computational technologies in journalism, we have drawn on research from the social sciences and neighboring disciplines that investigated data-intensive journalism. In combination, these two articles provide a functional roadmap for further developing and researching the field.

In line with the method of structured literature reviews, the first aim of this article is to *describe the development and current state of scholarship on data journalism*. This will help scholars assess both the development of the field and to identify works that have had the greatest influence on its current state. Consequently, our research questions are as follows:

RQ1. How is the research literature on data journalism developing, among other aspects, in terms of publication activity, publication outlets, and citations?

RQ2. What are the theories of data journalism research?

RQ3. What are the methods of data journalism research?

RQ4. Which research gaps have data journalism researchers identified and what are their suggestions for future research?

Each of these questions is answered in the 'Results' section of this article and lays the groundwork for the second aim of this article: *Providing propositions for future research on data journalism*. A reflection on the development, theories, methods, and research gaps in the 'Conclusion' section clarifies several possible areas of study, while noting that the means to explore them are in essence two: recombining established research interests, theories, and methods into novel approaches or integrating new interests, theories, or methods into existing frameworks. This article supports future research of either kind.

Data journalism and data-intensive newswork

Before proceeding, we would like to clarify what we mean by data journalism and data-intensive newswork – two umbrella terms used synonymously in this article. Both relate to computer-assisted reporting (CAR), a journalistic movement, which for decades has been concerned with the use of computational technologies in newsrooms (e.g. described in Howard, 2014; Stavelin, 2013). However, the focus with our term is to emphasize the increased availability of data and the technical sophistication of analysis as well as modes of representation, which have developed significantly since the beginning of CAR. As is often the case with emerging phenomena, there is no commonly agreed-upon definition of data journalism. Some definitions focus on the change in the process of newswork (Appelgren and Nygren, 2014b; Bounegru, 2012; Davenport et al., 2000; Diakopoulos, 2011; Felle, 2016; Flew et al., 2010; Karlsen and Stavelin, 2014; Parasie and Dagiral, 2013; Tandoc and Oh, 2015; Uskali and Kuutti, 2015; Weber and Rall, 2012; Weinacht and Spiller, 2014; Young and Hermida, 2015). These definitions stress the role of data as an additional source in the news-gathering process that requires special skills to handle. Other definitions emphasize that data journalism produces news items based on data analysis that include some form of interactive visualization, for example, maps or diagrams (Baack, 2013; Hullman et al., 2015). However, a large corpus of works highlights that data journalism is both a process and a product (Aitamurto et al., 2011; Appelgren and Nygren, 2014a; Ausserhofer, 2015; Coddington, 2015; Cohen et al., 2011a; Gynnild, 2014; Hamilton and Turner, 2009; Hannaford, 2015; Howard, 2014; Knight, 2015; Loosen et al., 2015; Radchenko and Sakoyan, 2014; Stavelin, 2013; Tabary et al., 2016). While each definition has a different emphasis, there are common elements which essentially outline a journalistic process (developing 'data stories' by analyzing large sets of data with (mostly) quantitative, computational methods), as well as a special form of presentation (interactive visualizations).[1] To capture this diversity, this article emphasizes the process and the product equally.

Prototypical examples include the *Guardian*'s 'Afghan' and 'Iraq War Logs' (mentioned in Baack, 2013, and Knight, 2015); 'ChicagoCrimes', a website which allowed users to check crimes in a particular neighborhood (mentioned in Parasie and Dagiral, 2013); and the 'Toxic Waters project' from the *New York Times* which examined pollution in American waters (mentioned in Gynnild, 2014). Such news items and the work leading to them have been described by a variety of terms (see Table 1), many of which are connected with different communities, histories, epistemologies, and visions of the public (Bounegru, 2012; Coddington, 2015). However, what is important to this article is what unites these practices: the concentrated use of structured information in the making of news. Thus, the terms data journalism and data-intensive newswork describe (the production of) news that is predominantly based on the concentrated collection and analysis of structured information. Like numerous other authors, we see this phenomenon as journalism's response to an increasingly data-dependent society.

Method

Methodological framework: Structured literature reviews

This article is a structured literature review, which represents a very rigid form of a meta-analysis (Massaro et al., 2016). The goal of a structured literature review is 'to
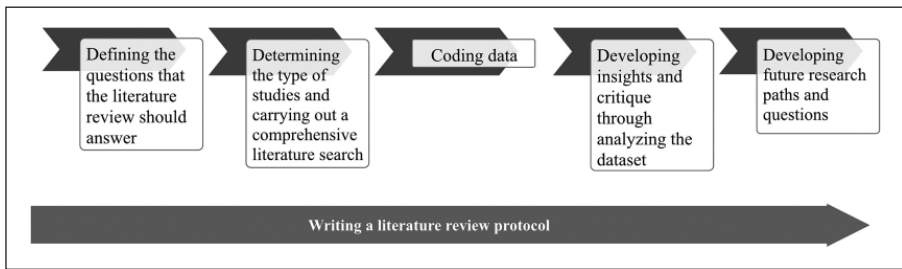
Figure 1. Undertaking a systematic literature review.
Source: Own visualization, based on Massaro et al. (2016).

develop insights, critical reflections, future research paths and research questions', and it should 'contribute to developing research paths and questions by providing a foundation on which to build on prior discoveries' (p. 768). In addition, structured literature reviews provide the basis and justification for new research, and the 'the background to develop research synthesis' (p. 794) for more mature areas of research. Systematic reviews adopt 'a replicable, scientific and transparent process […] that aims to minimize bias […]' (Tranfield et al., 2003: 209). While based on a 'positivist, quantitative, form-oriented content analysis method for reviewing literature', they also strongly use hermeneutic and interpretative methods, especially when developing insight and critique (Massaro et al., 2016: 768). The purpose of critique in this context 'is to counteract the dominance of taken-for granted goals, ideas, ideologies and discourses […]' (Alvesson and Deetz, 2000: 18).

Figure 1 illustrates the steps we took: First, we defined the questions the review should answer. Second, we selected relevant literature, following a rigid and reproducible path. Third, we analyzed and coded the material. In order to pay tribute to the shift toward more inclusive research findings in meta-analyses (Tsakalerou and Katsavounis, 2015), we integrated both qualitative and quantitative research.

Literature retrieval and selection

We started the process of literature retrieval by discussing formal inclusion and exclusion criteria for the literature corpus. These decisions were based on our evaluation of other structured literature reviews (e.g. Fecher et al., 2015; Guthrie and Murthy, 2009; Jungherr, 2016; Lecheler and Kruikemeier, 2016; Massaro et al., 2015; Neuberger et al., 2007). Although there are many excellent theoretical and conceptual publications on the subject, we decided to only include articles in the analysis that were empirically grounded, irrespective of whether the investigation focused on quantitative, qualitative, or mixed-methods research. This decision mirrored other meta-analyses and was made because we considered empirical proximity to be key for our research interest. Our focus was on publications from the social sciences, but we also included articles from other disciplines. In terms of publication type, we considered journal articles, book chapters, conference papers, industry reports, and PhD theses. Due to the high number of publications, we excluded bachelor's and master's theses, press reports, and blog posts, as well as popular textbooks such as the *Data Journalism Handbook* (Gray et al., 2012), *Precision*

Table 1. Synonymic search terms used for the literature retrieval.

Synonym	
Algorithmic journalism	Data journalism
Data-driven reporting	Data journalism
Computational journalism	Data-driven journalism
Database journalism	Quantitative journalism
Computer-assisted reporting	

Other related terms, more general concepts, and abbreviations such as 'accountability journalism', 'CAR', 'crowdsourced journalism', 'data visualization', 'DDJ', 'drone journalism', 'investigative journalism', 'online journalism', and 'open journalism' were tested as well but not included in the final list because the search results were too cluttered.

Journalism (Meyer, 2002), or *Scraping for Journalists* (Bradshaw, 2014) due to their target readership and self-reporting writing style. Due to limited language proficiency, we could not include publications written in languages other than in English or German. Finally, we decided to include all articles published online between 1996 and 2015, as the scattered research on CAR published before 1996 seemed to have investigated a very different practice. In total, two articles published in print in 2016 were accessible online during this evaluation period and were thus included in the review.

The next step was to create a list of relevant keywords that served as search terms for the literature retrieval from scientific databases. For this, we conducted a preliminary search on Google Scholar using the terms 'data-driven journalism' and 'data journalism' and then extracted related terms from the abstracts and keyword sections of the articles. This led us to the list of synonyms for data journalism as shown in Table 1.

Based on the synonymic search terms from Table 1, we undertook a literature search in 15 scientific databases, obtaining 772 records in total (Table 2). The search terms were combined with the Boolean 'OR' operator.

The records were imported into Zotero; assessed independently by two researchers considering title, abstract, and keywords; and selected according to the above-mentioned formal criteria and research focus (see first paragraph of this section for detailed criteria). Publications that both researchers marked as 'not relevant' were dismissed, while those both marked as 'relevant' were included in the preliminary corpus. The researchers discussed divergent assessments until an agreement was reached. This screening resulted in a preliminary corpus that we then cleaned of duplicates.

In total, two further measures were taken to select the literature for the corpus. First, we asked three domain experts, all researchers in the field of data journalism, to add relevant work based on their expertise. Second, we checked the references of our selection, computer supported and manually: We scraped the references of all selected articles from the preliminary corpus into a database, employing a bibliographic data recognition algorithm (Lopez, 2009) to assess all references that had been cited by at least two articles. Finally, the references of the selected works published in 2015 and 2016 were checked by hand. This process resulted in a corpus of 40 publications. Figure 2 illustrates the retrieval and selection process in a flowchart.

Table 2. Scientific databases used with the search terms from Table 1.

Database	Number of records		
ACM Digital	112	Sociological Abstracts	5
EBSCO	19	Sowiport	53
Google Scholar	400	Springer	29
IEEE	26	SpringerLink	135
JSTOR	33	Taylor & Francis Online	144
ProQuest	6	Web of Science	43
Science Direct	69	Wiley	73
Scopus	25		
		Total	772

Google Scholar provided us with 3290 results, but we only imported the first 400 because after that there were no more relevant hits.

Quantitative and qualitative content analyses

The final corpus of 40 research publications allowed us to perform different types of quantitative and qualitative analyses. We began by extracting and comparing different structural aspects of each piece. Among other dimensions, we collected the publishing date of each article and coded their publication outlets. Furthermore, we extracted the authors' names and affiliations, and the numbers of citations on Google Scholar using a Python module (Kreibich, 2016). Next, we took a closer look at the references, which we obtained using the same algorithm that we had employed for the preliminary corpus. We then cleaned the reference data, removed duplicates, and finally parsed it into a network-analysis-friendly format to identify central publications.

Parallel to the computational exploration of the corpus, we conducted a software-assisted qualitative content analysis (Kaefer et al., 2015; Mayring, 2000; Schreier, 2012), starting with the development of a codebook. While the codebook's main categories were defined by our research interest and research questions, we derived the subcategories out of the material manually. Such an inductive coding process helped to establish 'novel interpretative connections based on the data material, rather than [...] a conceptual pre-understanding' (Fecher et al., 2015: 6) of the topic. In a pre-test, three researchers coded two articles, continuously discussing and modifying the category system. We then compared the coding and discussed differences until an agreement on the codebook was reached. We used the qualitative data analysis software NVivo for the coding.

Finally, we supplemented the manual coding with an automated content analysis, which used a Python implementation of the term frequency-inverse document frequency (TF-IDF) algorithm to extract the central terms of each article. TF-IDF is a statistical measure that evaluates the importance of a word within an individual document, as well as the corpus as a whole.

Among other aspects, this mixed-method approach brought insights into the development of the research literature, the theoretical frameworks and the research designs used in the publications, and the areas potentially neglected within the field – matters addressed in the following section.
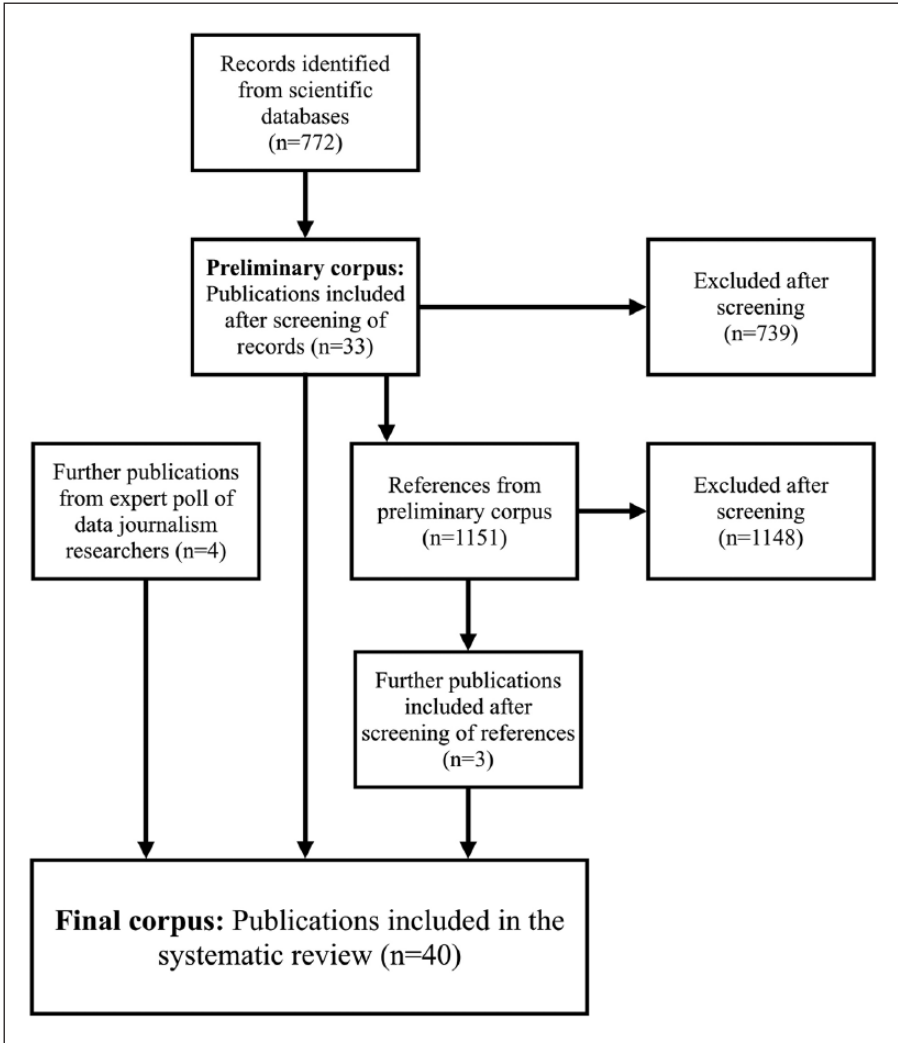
Figure 2. Workflow of literature retrieval and selection.
Source: Own visualization, adapted from Fecher et al. (2015: 2).

Results

The results presented here refer to the research questions formulated in the 'Introduction' section. In the first subsection, we go into detail about the current state of data journalism research and provide a citation and collaboration analysis. The following two subsections explore the employed theoretical frameworks and research methods. The final subsection covers the research gaps identified by other scholars. For further perspectives on the corpus, please also visit our interactive literature explorer at http://literature.validproject.at.
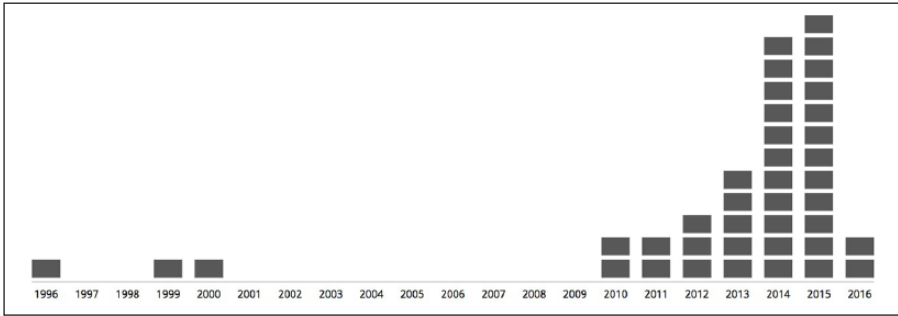
Figure 3. Development of the literature over time.
Every block represents one publication from the corpus. The two articles from 2016 were published online before print within our evaluation period and could, therefore, be included in the review.

Corpus characteristics and bibliometry

Our first research question was: How is the research literature on data journalism developing? Figure 3 indicates that there has been an increase of research publications on data journalism and related fields since 2010. Although what is known as CAR has been practiced since the 1960s (Cox, 2000), the scientific investigation of it has started only recently. Before 2010, only a small number of isolated research publications dealt with the topic of data and journalism. Authored by US journalism researchers, these first publications mainly investigated how journalists used different computational technologies in the newsroom (Davenport et al., 1996, 2000; Garrison, 1999). The recent increase of research activity on the subject has to be seen in association with the introduction or renaming of different practices of data-intensive newswork as 'data journalism' or 'data-driven journalism' in the late 2000s. This reframing seemed to have sparked the interest of journalism practitioners and journalism researchers alike.

Academic journals focused on journalism research published the largest number of works (26) in the corpus. Among them the periodicals *Digital Journalism* (6), *Journalism Studies* (3), *Journalism* (2), and *Journalism Practice* (2) were the most popular places to publish. *Digital Journalism* owes its lead position to the publication of a special issue (Lewis, 2015). Journal articles not only outnumber other channels of academic dissemination, they were also cited considerably more often than conference articles (6), book chapters (4), and other publication types (4; see Figure 4). The most-cited article in the corpus (Segel and Heer, 2010) was published in a computer science journal.

What are the most influential publications in the field of data journalism research thus far? A count of the references of all publications hinted at an answer, though a high number of citations does not necessarily equal high influence, as many cited publications are not geared toward an academic audience. Phil Meyer's (2002) book *Precision Journalism*, which has helped to build the 'computer-assisted reporting legacy' (Parasie and Dagiral, 2013) since the 1970s, was the most-cited work in the corpus together with Parasie and Dagiral's investigation of the community and epistemologies of data journalists in Chicago. Their article can be considered as a prototypical piece of contemporary data journalism research since its theoretical framework, research methods, and scope of investigation have

resurfaced in many later publications. Both publications, Meyer's and Parasie and Dagiral's, have been cited by 15 of the 40 articles in our corpus. Other often-cited and influential works on data journalism can be seen as part of research that has 'focused on the techno-logical *promises* of computing on journalism' (Gynnild, 2014: 714) or publications that explore the potential of the practice (Cohen et al., 2011a, 2011b; Flew et al., 2012; Hamilton and Turner, 2009). Table 3 lists the 10 most-cited publications in our corpus.
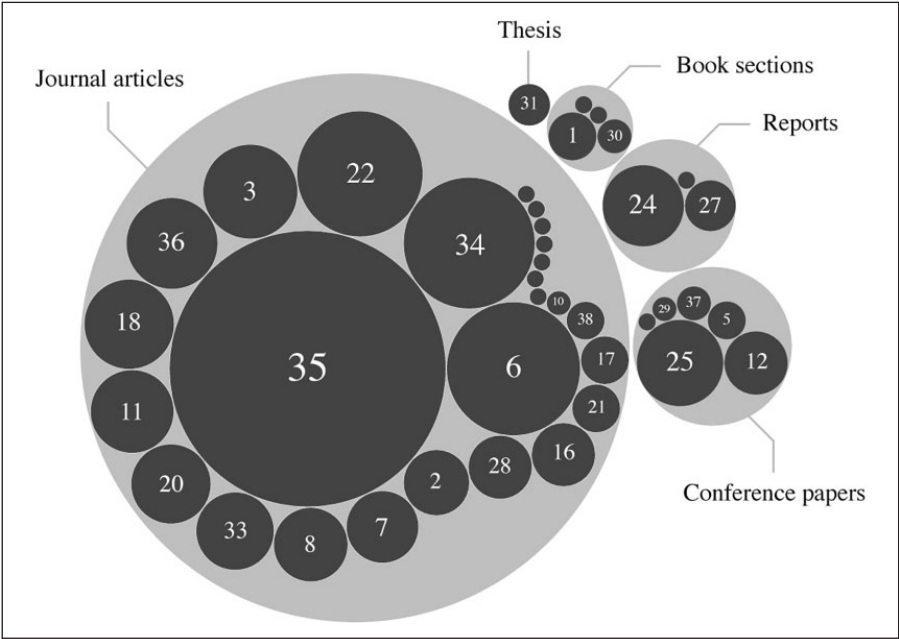

Figure 4. Publications by type and number of citations.
The numbers can be referenced in Table 4. Diameter of bubble = number of citations on Google Scholar.
Last revision of Google Scholar stats on 12 April 2016.

Table 3. The most-cited references.

Publication	Number of citations
Meyer (2002)	15
Parasie and Dagiral (2013)	15
Gray et al. (2012)	13
Flew et al. (2012)	11
Hamilton and Turner (2009)	10
Royal (2012)	10
Cohen et al. (2011a)	8
Cox (2000)	7
Cohen et al. (2011b)	7
Lewis and Usher (2013)	6

Where was the research on data journalism produced? Geo-coding the first affiliation of each author as they were mentioned in a publication, we found that most researchers came from institutions in Europe (38) and North America (29). Of the 71 authors in our corpus, 22 were affiliated with institutions in the United States. Clearly a collaborative field, more than two-thirds (27) of all data journalism publications were written by more than one author. For comparison, in 2000, roughly half of social science articles were written collaboratively (Wuchty et al., 2007). In total, seven of the articles have been international collaborations. Figure 5 places the authors' affiliations on a map and shows that almost no non-Western research on data journalism has been published in English.[2]

Theoretical frameworks

Journalism studies in general tends to be a 'multidisciplinary field in the borderland between humanities, social science and technology' (Appelgren and Nygren, 2014a). Hence, several articles in the corpus mentioned certain theoretical frameworks which are to be understood in a broader sense than when applied in their original fields. These articles also revealed how research on data journalism situated itself within the research stream, and as always, theoretical concepts had an influence on the selection of possible objects of research, on what to investigate specifically, and finally which research methodologies were selected and made explicit.

Theoretical frameworks were often combined, suggesting that certain clusters of frameworks were frequently used in data journalism studies. Science and technology *studies* were repeatedly mentioned (Ausserhofer, 2015; Lewis and Usher, 2013, 2014; Stavelin, 2013), and actor network theory was often applied in the analysis (Ausserhofer, 2015; De Maeyer et al., 2015; Parasie, 2015; Parasie and Dagiral, 2013; Stavelin, 2013). More specific theoretical concepts were also used, for example, 'understanding […] computational journalism as a rhetorical craft, using the Aristotelian concept of techné' (Karlsen and Stavelin, 2014), role theory (Weinacht and Spiller, 2014), and trading zones (Lewis and Usher, 2014). The fact that only few publications mentioned theoretical frameworks might indicate that data journalism research is more oriented toward practice than theory currently, which is very common for journalism studies (Scholl, 2011).

Research designs

The analysis of research designs is summarized in Table 4. It revealed a certain scarcity of quantitative research designs and digital methods, in contrast to qualitative exploratory ones. At this early stage of research, this is not unusual since the characteristics of a new practice have to be explored and defined before quantitative methods can be applied.

Qualitative interviews were by far the most common method (25 publications) used within the examined literature corpus. Many of them were in-depth interviews that followed semi-structured guidelines and were conducted with practitioners and/or experts ranging from 5 to 35 individuals or in 1 case over 100 (Howard, 2014). In total, seven of the interviews questioned 5 to 9 individuals, six interviewed between 10 and 19, and six interviewed over 20 individuals. Interviews either took place in person or via phone. Commonly, those interviews were the main method used for the publication, though
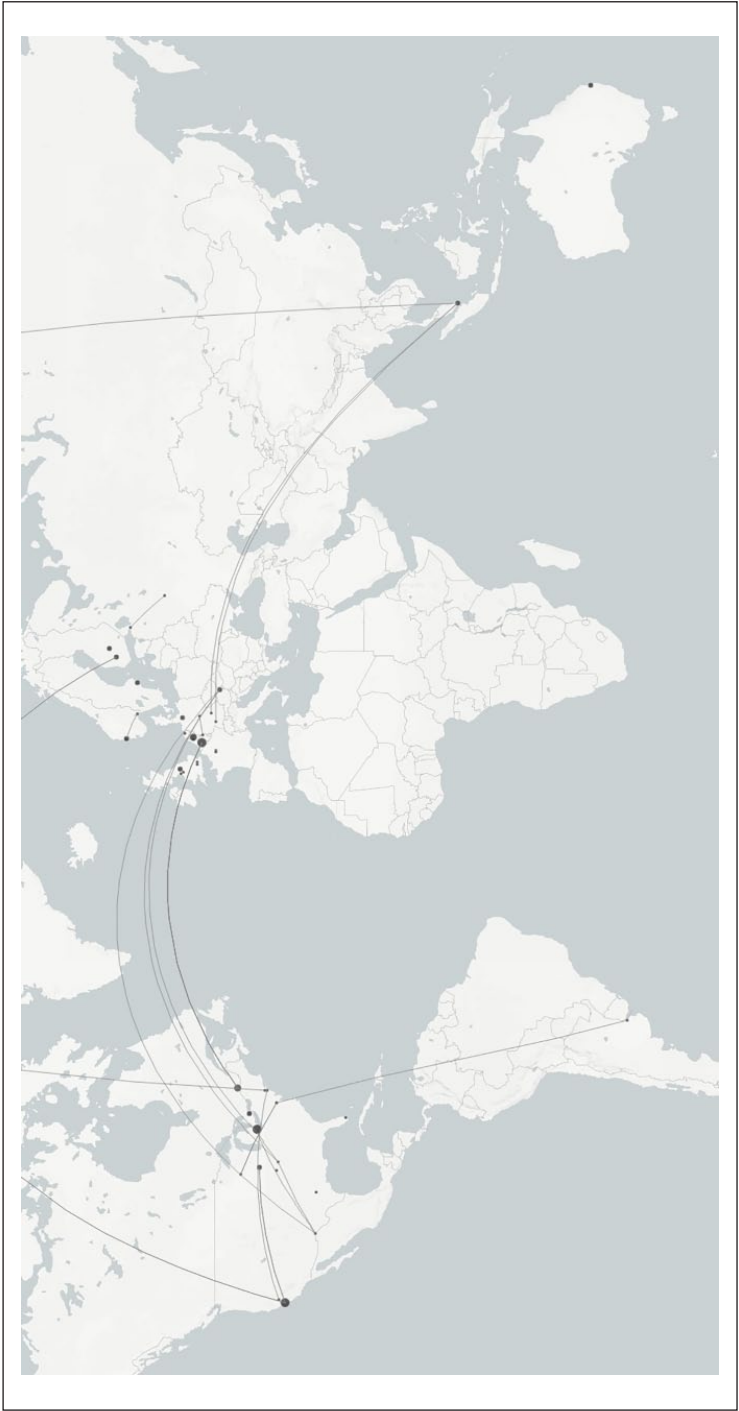
Figure 5. Affiliations and collaborations.
Larger dots are indicators of several authors at one institute.

Table 4. Research methods of publications on data journalism.

ID	Publication	QI	CA	Other method	Geographical focus (country)	Topic
1	Baack (2013)			Case study	N/A	*Wikileaks* and DIN
2	Lewis and Usher (2014)			Short-term observation	International	Case study of Hacks/Hackers
3	Karlsen and Stavelin (2014)				Norway	DIN in Norway
4	Hannaford (2015)				United Kingdom	DIN at the *BBC* and the *Financial Times*
5	Hullman et al. (2015)				United Kingdom	User comments at the *Economist*'s graphic detail blog
6	Parasie and Dagiral (2013)				United States	DIN at the *Chicago Tribune*
7	Parasie (2015)				United States	Epistemology of DIN
8	Appelgren and Nygren (2014b)			Survey	Sweden	DIN at Swedish media organizations
9	Appelgren and Nygren (2014a)			Survey	Sweden	DIN at Swedish media organizations
10	Knight (2015)				United Kingdom	DIN of daily and Sunday newspapers
11	Fink and Anderson (2015)				United States	DIN at small, medium, and large newspapers in the United States
12	Weber and Rall (2012)				United States, Germany, Switzerland	DIN at different online newspapers
13	Weinacht and Spiller (2014)				Germany	Self-perception of German data journalists
14	Ausserhofer (2015)				Austria, Germany, United Kingdom	Editorial workflows
15	Felle (2016)				International	DIN and the 'fourth estate'
16	Young and Hermida (2015)			Short-term observation	United States	Crime journalism at the *LA Times*
17	Dick (2014)			Short-term observation	United Kingdom	Infographics in UK online news
18	Gynnild (2014)				United Kingdom, United States	Journalism innovation and its rhetoric
19	Uskali and Kuutti (2015)				Finland, United Kingdom, United States	DIN practices in different newsrooms
20	Bakker (2014)				Netherlands	Changing roles of journalists
21	Garrison (1999)			Survey	United States	DIN resources of daily US newspapers
22	Lewis and Usher (2013)			Close reading	N/A	Open source in journalism
23	Lugo-Ocando and Brandão (2015)				United Kingdom	Crime-reporting and statistics
24	Howard (2014)				International	State of DIN in general

(Continued)

Table 4. (Continued)

ID	Publication	QI	CA	Other method	Geographical focus (country)	Topic
25	Daniel and Flew (2010)		■		United Kingdom	The *Guardian*'s coverage of a UK expenses scandal
26	Radchenko and Sakoyan (2014)		■		Russia	Open data and DIN in Russia
27	Aitamurto et al. (2011)	■			United Kingdom, United States, Argentina	State of DIN in general
28	De Maeyer et al. (2015)	■			(French-speaking) Belgium	Regional DIN and its discourse
29	Loosen et al. (2015)		■		International	Pieces nominated for the International Data Journalism Awards
30	Weber and Rall (2013)	■			United States, Germany, Switzerland	DIN at different online newspapers
31	Stavelin (2013)	■			Norway	DIN in Norway
32	Tandoc and Oh (2015)	■			United Kingdom	News values, norms, and routines of the *Guardian Datablog*
33	Royal (2012)			Newsroom ethnography	United States	DIN at the *New York Times*
34	Cohen et al. (2011a)	■			N/A	Impact of DIN on accountability reporting
35	Segel and Heer (2010)		■		United Kingdom, United States	Narrative visualizations in online journalism and other fields
36	Davenport et al. (1996)			Survey	United States	Computer use in Michigan newsrooms
37	Davenport et al. (2000)			Survey	United States	DIN in Michigan's newspapers
38	Smit et al. (2014)	■	■	Short-term observation	Netherlands	Production of news visualization
39	Zanchelli and Crucianelli (2012)	■	■		United Kingdom, United States, Brazil	DIN processes in newsrooms in the United States, the United Kingdom, and Brazil
40	Tabary et al. (2016)	■	■		Canada	DIN in Quebec

DIN: data-intensive newswork. QI: Qualitative interviews. CA: Content analysis.
Content analysis includes analysis of news, databases, blogs, job ads, visualizations, briefings, manuals, and more. Short-term observation encompasses visits to the newsroom and participation in meetings. A newsroom ethnography is defined as an observational study of a newsroom over the course of several days.

many publications employed mixed-method approaches. In total, eight used interviews in addition to other methods. In some cases, initial interviews were supported by content or document analysis and vice versa. For instance, De Maeyer et al. (2015) conducted 20 semi-structured interviews with individuals involved in data journalism in Belgium, while also analyzing documents and artifacts.

Content analysis was also employed quite often. Overall, we were able to identify 21 publications within our corpus that use some sort of analysis of text or images, though very few actually examined data journalistic news items. Examples of these included an article by Lugo-Ocando and Brandão (2015) that analyzed news items produced by journalists in the United Kingdom, and one publication, which analyzed articles within the *Guardian Datablog* with respect to news values, sources, and topics (Tandoc and Oh, 2015). Loosen et al. (2015) focused on pieces nominated for the Data Journalism Awards in the years 2013 and 2014. Segel and Heer (2010) chose to explore 58 visualization examples from online media.

Other research publications employing content analysis looked at the discourse around data journalism and not data journalistic news items themselves. For instance, Hullman et al. (2015) analyzed comments from the *Economist*'s *Graphic Detail* blog, while Gynnild (2014) studied both original data journalistic items – pieces on the *Guardian Datablog* – and online news accounts, listservs, and journalism blogs to grasp the discourse around data journalism.

The method of survey was used less frequently, though Appelgren and Nygren (2012a, 2014b) made use of an online survey and combined the findings with semi-structured interviews for their research.

Some authors used observation as a method to study newsrooms, though in the case of Royal (2012), who observed members of the *New York Times*' Interactive News Technology department for several days, it seemed appropriate to classify her method as ethnography. Smit et al. (2014) attended editorial meetings and brainstorming sessions at a leading broadcasting organization in the Netherlands, while Dick (2014) spent 8 hours observing the *BBC* News Online Specials team.

The studied cases tended to be in the United States and the United Kingdom, though some were also in Western countries such as Germany and Sweden. There exists almost no English language research on practices outside of Europe and North America. Some author collectives tended to focus on activities in less explored countries, for instance, Sweden (Appelgren and Nygren, 2014a, 2014b) and Norway (Karlsen and Stavelin, 2014; Stavelin, 2013). The restriction to certain geographical areas coincides largely with the researchers' location. The fact that even large Western countries such as France are not represented in our sample might be partly due to the lack of a longstanding tradition of CAR there (Parasie and Dagiral, 2013).

Which media organizations did researchers associate with data journalism and investigate? The list in Figure 6 was compiled through manual coding in combination with automatic term extraction of the texts. The visualization shows the frequency and total mentions of a specific media organization, as well as how many publications were connected to it. The *Guardian*, the *New York Times*, and *ProPublica* occurred most often and in most documents, but Figure 6 also helps to identify smaller news organizations that were only rarely mentioned but might be suitable for future research.

Research gaps

Given the emerging nature of the field, we wanted to know what scholars identified as research gaps. Recommendations for further study included cross-national investigations
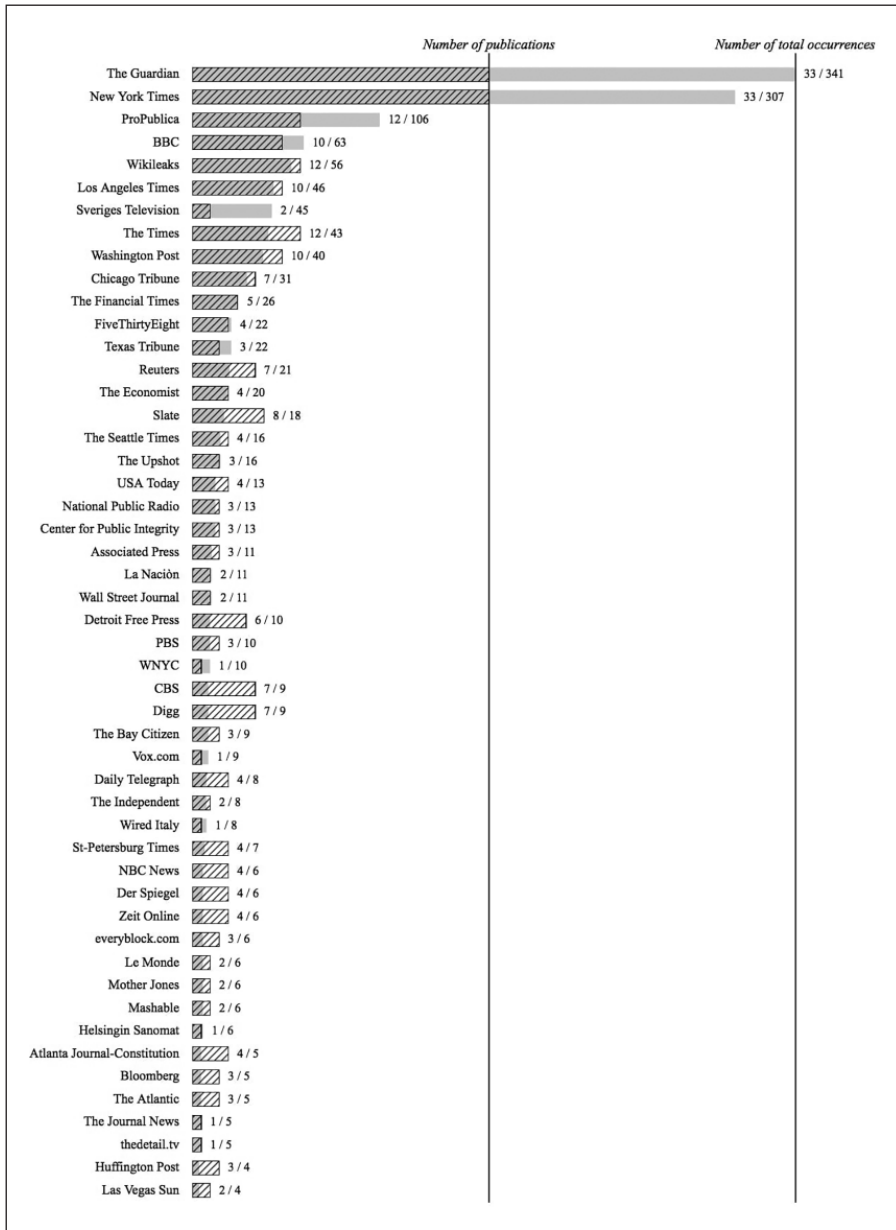
Figure 6. Occurrence of media organizations.

of data journalism, as well as the ethnographic study of those practices. Parasie and Dagiral (2013), for example, suggested comparing practices between countries while accounting for the differences in the cultures of journalists and hackers. They suggested that those differences influenced the way data journalism is practiced within various

countries. They also proposed that ethnographic studies of newsrooms could clarify how programmer–journalists were integrated into different organizations. Along those lines, Appelgren and Nygren (2014b) suggested further comparative international research that took national regulations and constraints into account.

Most of the publications we analyzed focused on a rather short period. Therefore, Knight (2015) pointed out the need for long-term studies, which Davenport et al. (2000) had already recommended more than a decade before. Interested in the influence of technology on content, Garrison (1999) recommended determining whether the discrepancy of resources between small and large newspapers led them to tell different stories (scope, depth, and size of databases). Lewis and Usher (2013) also suggested research into the influence of technology on the news-production process, while Stavelin (2013) more specifically called for the study of how software design and its use affected the production of news. Segel and Heer (2010), however, proposed studies focused on reader experience to clarify how audiences engaged with visual elements and, therefore, how their design could be improved. Despite this variety of suggestions, most authors and author collectives within our corpus did not discuss research gaps or suggest paths for further research.

Conclusion

In this article, we analyzed the structure and topics of research literature on data journalism that has been published in the last 20 years. Following the rigid method of a structured literature review, we selected a corpus of 40 publications and examined them using computational methods and qualitative software-assisted content analysis.

Both data journalism and its accompanying research have been developing rapidly in the past two decades. Since 2014 in particular, we have seen a major increase of scholarship on the topic. This growth has led to quality improvements and contributed to the establishment of a solid foundation for the field. An indicator for the quality improvement is that newer publications increasingly reference publications that have been produced by researchers following consistent methods and published in peer-reviewed journals. Another indicator for improvement is the high percentage of collaborations within the research community. This coincides with the 'collaboration imperative' of modern research which is also found to have beneficial impacts regarding the creation and diffusion of knowledge (Bozeman and Boardman, 2014). The frequent collaboration across countries can be read as a tendency toward internationalization.

At the same time, we also see issues with parts of the current literature. For instance, only a minority of empirical data journalism research refers to theoretical or methodological concepts. Many publications just report what has been investigated. Thus, we welcome research that either draws from different strands of theory or continues to develop theoretical concepts. This effort could build upon those of a number of scholars doing theoretical work related to the role of data and journalism in our society (e.g. Anderson, 2015; Bunz, 2011; Cohen, 2015; Fairfield and Shtein, 2014; Lewis and Westlund, 2015; Schudson, 2010). For instance, capturing concepts of data journalism through a history-of-ideas lens could be very enriching. While we acknowledge that

descriptive research is very important – especially with new phenomena – it also seems like some data journalism researchers have adopted the objectivist epistemology of many journalists working with quantitative data (Godler and Reich, 2013). This means, for example, that some authors reported their results without taking their method of data collection into account: responses of interviewees are presented as facts without acknowledging that some degree of idealization is inevitable. In most interview-based publications, the respondents were personally identifiable; given the fact that many of the practitioners engage in extra-institutional communities (Usher, 2016), which monitor and shape the field's (meta) discourse, it can be assumed that the interviewees tried to present themselves in the best possible way.

Another issue is the lack in variety of research methods. While many research projects employ two different qualitative methods or triangulation, there is almost no research based on quantitative methods. While exploratory studies are typical for newly developing research fields, it seems appropriate to begin testing some existing theories using larger samples. Furthermore, we believe that 'digital methods' (Rogers, 2013; see also Maireder et al., 2015), or methods that try to grasp the field through online trace data or by observing the interactions on the major platforms for data journalism discourse, such as *GitHub*, *Slack*, *Twitter*, *Facebook*, or *Meetup*, would offer new avenues for data journalism research.

In addition to the research gaps described above, data journalism research currently has little to say about questions of gender. As women seem to be a minority in data journalism, some articles proposed gender issues regarding the (male-dominated) field of technology as general topic. However, questions concerning the more specific interests, experiences, or challenges of women in data journalism were rudimentarily addressed. We would be happy to see research focusing on this and similar matters.

Finally, very little is known about data journalism outside of the news desks of famous organizations. We would thus welcome research directed toward local and mobile data journalism, toward small news outlets working on data-intensive projects, and toward the many freelancers who provide services for news organizations. Here, an economic perspective would be an especially valuable contribution to the field.

In journalism research, there is a strong connection between the research interest, the chosen theory, the methods of data collection and analysis, and the reporting of results (Scholl, 2011). Some works combine these aspects in a systematic way, while others generate new perspectives from integrating new or remixing proven perspectives, theories, and methods into novel frameworks (Markham, 2013). Whichever way is chosen for future investigations on data journalism – a traditional or a bricolage approach – this article provides the groundwork for both. By critically surveying essential elements of publications and providing additional research propositions, it allows future researchers to choose their research interests, theoretical concepts, and methods with respect to the continuity and innovation of the field of data journalism research.

Acknowledgements

Joseph Robinson, Elena Rudkowsky, Michael Sedlmair, Thomas Wolkinger, and the two anonymous reviewers for their helpful feedback.

Notes

1. Practices involving 'simpler' forms of data representations for journalistic purposes (e.g. static infographics) are not in the focus of this research.
2. During the literature retrieval phase, we identified some publishing activity in other languages, for instance, in French ('journalisme de données') and Spanish ('jornalismo de dados'). These publications could not be reviewed, see the respective subsection for selection limitations.

References

Aitamurto T, Sirkkunen E and Lehtonen P (2011) Trends in data journalism. Available at: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf

Alvesson M and Deetz S (2000) *Doing Critical Management Research*. London: SAGE.

Anderson CW (2013) Towards a sociology of computational and algorithmic journalism. *New Media & Society* 15(7): 1005–1021.

Anderson CW (2015) Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism. *Digital Journalism* 3(3): 349–363.

Appelgren E and Nygren G (2014a) Data journalism in Sweden – Opportunities and challenges: A case study of Brottspejl at Sveriges Television (SVT). In: Mills C, Pidd M and Ward E (eds) *Proceedings of the Digital Humanities Congress 2012*. Sheffield: HRI Online Publications. Available at: https://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren

Appelgren E and Nygren G (2014b) Data journalism in Sweden: Introducing new methods and genres of journalism into 'old' organizations. *Digital Journalism* 2(3): 394–405.

Ausserhofer J (2015) 'Die Methode liegt im Code': Routinen und digitale Methoden im Datenjournalismus. In: Maireder A, Ausserhofer J, Schumann C and et al. (eds) *Digitale Methoden in der Kommunikationswissenschaft*. Berlin: Digital Communication Research, pp. 87–111. Available at: http://dx.doi.org/10.17174/dcr.v2.5

Baack S (2013) A new style of news reporting: Wikileaks and data-driven journalism. In: Rambatan B and Johanssen J (eds) *Cyborg Subjects: Discourses on Digital Culture*. CreateSpace Independent Publishing, pp. 113–122.

Bakker P (2014) Mr. Gates returns: Curation, community management and other new roles for journalists. *Journalism Studies* 15(5): 596–606.

Bounegru L (2012) Data journalism in perspective. In: Gray J, Bounegru L and Chambers L (eds) *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Sebastopol, CA: O'Reilly, pp. 17–22.

Bozeman B and Boardman C (2014) *Research Collaboration and Team Science: A State-of-the-Art Review and Agenda*. Cham: Springer.

Bradshaw P (2014) *Scraping for Journalists*. Leanpub.

Bunz M (2011) Das offene Geheimnis: Zur Politik der Wahrheit im Datenjournalismus. In: Geiselberger H (ed.), *Wikileaks und die Folgen: Netz - Medien - Politik*. Berlin: Suhrkamp, pp. 134–151.

Coddington M (2015) Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism* 3(3): 331–348.

Cohen NS (2015) From pink slips to pink slime: Transforming media labor in a digital age. *The Communication Review* 18(2): 98–122.

Cohen S, Hamilton JT and Turner F (2011a) Computational journalism: How computer scientists can empower journalists, democracy's watchdogs, in the production of news in the public interest. *Communications of the ACM* 54(10): 66–71.

Cohen S, Li C, Yang J, et al. (2011b) Computational journalism: A call to arms to database researchers. In: *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research*, Asilomar, CA, 9–12 January, pp. 148–151. Available at: http://cidrdb.org/cidr2011/Papers/CIDR11_Paper17.pdf

Cox M (2000) The development of computer-assisted reporting. In: *Proceedings of the Newspaper Division of the Association for Education in Journalism and Mass Communication's Southeast Colloquium*, Chapel Hill, NC, 17–18 March.

Daniel A and Flew T (2010) The Guardian reportage of the UK MP expenses scandal: A case study of computational journalism. In: Papandrea F and Armstrong M (eds) *Record of the Communications Policy and Research Forum 2010*. Sydney, NSW, Australia: Network Insight, pp. 186–194.

Davenport L, Fico F and Detwiler M (2000) Computer–assisted reporting in Michigan daily newspapers: More than a decade of adoption. In: *Proceedings of the national convention of the association for education in journalism and mass communication*, Phoenix, AZ, 9–12 August.

Davenport L, Fico F and Weinstock D (1996) Computers in newsrooms of Michigan's newspapers. *Newspaper Research Journal* 17(3–4): 14–28.

De Maeyer J, Libert M, Domingo D, et al. (2015) Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432–446.

Diakopoulos N (2011) A functional roadmap for innovation in computational journalism. Available at: http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/

Diakopoulos N (2012) Cultivating the landscape of innovation in computational journalism. CUNY Graduate School of Journalism & Tow-Knight Center for Entrepreneurial Journalism. Available at: http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos_whitepaper_systematicinnovation.pdf

Dick M (2014) Interactive infographics and news values. *Digital Journalism* 2(4): 490–506.

European Journalism Centre (ed.) (2010) *Data-Driven Journalism: What Is There to Learn?* Amsterdam: European Journalism Center. Available at: http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf

Fairfield J and Shtein H (2014) Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics* 29(1): 38–51.

Fecher B, Friesike S and Hebing M (2015) What drives academic data sharing? *PLoS ONE* 10(2): e0118053.

Felle T (2016) Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism* 17(1): 85–96.

Fink K and Anderson CW (2015) Data journalism in the United States: Beyond the 'usual suspects'. *Journalism Studies* 16(4): 467–481.

Flew T, Daniel A and Spurgeon CL (2010) The promise of computational journalism. In: *Media, democracy and change: Refereed proceedings of the Australian and New Zealand Communication Association annual conference* (ed K McCallum), Canberra, ACT, Australia, 7–9 July.

Flew T, Spurgeon C, Daniel A, et al. (2012) The promise of computational journalism. *Journalism Practice* 6(2): 157–171.

Garrison B (1999) Newspaper size as a factor in use of computer-assisted reporting. *Newspaper Research Journal* 20(3): 72-85.

Godler Y and Reich Z (2013) How journalists think about facts. *Journalism Studies* 14(1): 94–112.

Gray J, Bounegru L and Chambers L (eds) (2012) *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Sebastopol, CA: O'Reilly.

Guthrie G and Murthy V (2009) Past, present and possible future developments in human capital accounting: A tribute to Jan-Erik Gröjer. *Journal of Human Resource Costing & Accounting* 13(2): 125–142.

Gynnild A (2014) Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism* 15(6): 713–730.

Hamilton JT and Turner F (2009) Accountability through algorithm: Developing the field of computational journalism. In: Proceedings of the center for advanced study in the behavioral sciences summer workshop, Palo Alto, CA. Available at: http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Hannaford L (2015) Computational journalism in the UK newsroom: Hybrids or specialists? *Journalism Education* 4(1): 6–21.

Howard AB (2014) *The Art and Science of Data-Driven Journalism*. New York: Tow Center for Digital Journalism. Available at: http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf

Hullman J, Diakopoulos N, Momeni E, et al. (2015) Content, context, and critique: Commenting on a data visualization blog. In: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, Vancouver, BC, Canada, 14–18 March, pp. 1170–1175. New York: ACM.

Jungherr A (2016) Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics* 13(1): 72–91.

Kaefer F, Roper J and Sinha P (2015) A software-assisted qualitative content analysis of news articles: Example and reflections. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 16(2). Available at: http://www.qualitative-research.net/index.php/fqs/article/view/2123

Karlsen J and Stavelin E (2014) Computational journalism in Norwegian newsrooms. *Journalism Practice* 8(1): 34–48.

Knight M (2015) Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice* 16(1): 55–72.

Kreibich C (2016) scholar.py. Available at: https://github.com/ckreibich/scholar.py

Lecheler S and Kruikemeier S (2016) Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources. *New Media & Society* 18(1): 156–171.

Lewis SC (2015) Journalism in an era of big data. *Digital Journalism* 3(3): 321–330.

Lewis SC and Usher N (2013) Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society* 35(5): 602–619.

Lewis SC and Usher N (2014) Code, collaboration, and the future of journalism: A case study of the hacks/hackers global network. *Digital Journalism* 2(3): 383–393.

Lewis SC and Westlund O (2015) Big data and journalism. *Digital Journalism* 3(3): 447–466.

Loosen W, Reimer J and Schmidt F (2015) When data become news: A content analysis of data journalism pieces. In: *Proceedings of the future of journalism 2015 conference*, Cardiff, 10–11 September.

Lopez P (2009) GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In: Agosti M, Borbinha J, Kapidakis S, et al. (eds) *Research and Advanced Technology for Digital Libraries*. Berlin: Springer, pp. 473–474.

Lugo-Ocando J and Brandão RF (2015) Stabbing news: Articulating crime statistics in the newsroom. *Journalism Practice* 10: 715–726.

Maireder A, Ausserhofer J, Schumann C, et al. (eds) (2015) *Digitale Methoden in der Kommunikationswissenschaft*. Berlin: Digital Communication Research. Available at: https://doi.org/10.17174/dcr.v2.0

Markham A (2013) Remix cultures, remix methods: Reframing qualitative inquiry for social media contexts. In: Denzin NK and Giardina MD (eds) *Global Dimensions of Qualitative Inquiry*. Walnut Creek, CA: Left Coast Press, pp. 63–81.

Massaro M, Dumay J and Garlatti A (2015) Public sector knowledge management: A structured literature review. *Journal of Knowledge Management* 19(3): 530–558.

Massaro M, Dumay J and Guthrie J (2016) On the shoulders of giants: Undertaking a structured literature review in accounting. *Accounting, Auditing & Accountability Journal* 29: 767–801.

Mayring P (2000) Qualitative content analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 1(2). Available at: http://www.qualitative-research.net/index.php/fqs/article/view/1089

Meyer P (2002) *Precision Journalism: A Reporter's Introduction to Social Science Methods*, 4th edn. Oxford: Rowman & Littlefield.

Neuberger C, Nuernbergk C and Rischke M (2007) Weblogs und Journalismus: Konkurrenz, Ergänzung oder Integration? *Media Perspektiven* 2007(2): 96–112.

Parasie S (2015) Data-driven revelation? Epistemological tensions in investigative journalism in the age of 'big data'. *Digital Journalism* 3(3): 364–380.

Parasie S and Dagiral E (2013) Data-driven journalism and the public good: 'Computer-assisted-reporters' and 'programmer-journalists' in Chicago. *New Media & Society* 15(6): 853–871.

Radchenko I and Sakoyan A (2014) The view on open data and data journalism: Cases, educational resources and current trends. In: Ignatov DI, Khachay MY, Panchenko A, et al. (eds) *Analysis of Images, Social Networks and Texts*. Cham: Springer, pp. 47–54.

Rogers R (2013) *Digital Methods*. Cambridge, MA: MIT Press.

Rogers S (2008) Turning official figures into understandable graphics, at the press of a button. *The Guardian*, December. Available at: http://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata

Royal C (2012) The journalist as programmer: A case study of the New York Times Interactive News Technology department. *ISOJ Journal* 2(1).

Scholl A (2011) Der unauflösbare Zusammenhang von Fragestellung, Theorie und Methode. In: Jandura O, Quandt T and Vogelgesang J (eds) *Methoden der Journalismusforschung*. Wiesbaden: Springer, pp. 15–32.

Schreier M (2012) *Qualitative Content Analysis in Practice*. Thousand Oaks, CA: SAGE.

Schudson M (2005) Four approaches to the sociology of news. In: Curran J and Gurevitch M (eds) *Mass Media and Society*, 4th edn. London: Hodder Arnold, pp. 172–197.

Schudson M (2010) Political observatories, databases & news in the emerging ecology of public information. *Daedalus* 139(2): 100–109.

Segel E and Heer J (2010) Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics* 16(6): 1139–1148.

Smit G, de Haan Y and Buijs L (2014) Visualizing news: Make it work. *Digital Journalism* 2(3): 344–354.

Stavelin E (2013) *Computational journalism: When journalism meets programming*. PhD Dissertation, University of Bergen, Bergen.

Steensen S and Ahva L (2015) Theories of journalism in a digital age. *Journalism Practice* 9(1): 1–18.

Tabary C, Provost A-M and Trottier A (2016) Data journalism's actors, practices and skills: A case study from Quebec. *Journalism* 17(1): 66–84.

Tandoc EC Jr and Oh S-K (2015) Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism Studies*. Epub ahead of print 5 November. DOI: 10.1080/1461670X.2015.1104260.

Tranfield D, Denyer D and Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14(3): 207–222.

Tsakalerou M and Katsavounis S (2015) Systematic reviews and metastudies: A meta-analysis framework. *International Journal of Science and Advanced Technology* 5(1): 23–26.

Usher N (2016) *Interactive Journalism: Hackers, Data, and Code*. Urbana, IL: University of Illinois Press.

Uskali T and Appelgren E (2015) The third Nordic data journalism conference: Call for papers. Available at: http://www.noda2016.fi/callforpapers.html

Uskali T and Kuutti H (2015) Models and streams of data journalism. *The Journal of Media Innovations* 2(1): 77–88.

Weber W and Rall H (2012) Data visualization in online journalism and its implications for the production process. In: *Proceedings of the 16th international conference on information visualisation (IV)*, 11–13 July, pp. 349–356. Montpellier: IEEE.

Weber W and Rall H (2013) 'We are journalists': Production practices, attitudes and a case study of the New York Times newsroom. In: Weber W, Burmester M and Tille R (eds) *Interaktive Infografiken*. Berlin: Springer, pp. 161–172.

Weinacht S and Spiller R (2014) Datenjournalismus in Deutschland: Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik* 59(4): 411–433.

Wuchty S, Jones BF and Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827): 1036–1039.

Young ML and Hermida A (2015) From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. *Digital Journalism* 3(3): 381–397.

Zanchelli M and Crucianelli S (2012) Integrating data journalism into newsrooms. *International Center for Journalists*. Available at: http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english_0.pdf

Author biographies

Julian Ausserhofer is an external PhD candidate in communication at the University of Vienna and an associate researcher at the Alexander von Humboldt Institute for Internet and Society in Berlin. His areas of interest include computational methods in the social sciences and in journalism and social media in news and politics.

Robert Gutounig is a research fellow and a member of the academic staff at the Department of Media and Design at the FH JOANNEUM University of Applied Sciences in Graz, Austria. His research focuses on media transformations in the digital age.

Michael Oppermann is a master's student in computer science at the University of Vienna. His research interests are human–computer interaction, visualization, visual analytics, and text mining.

Sarah Matiasek is a master's student in communication at the University of Vienna and a research assistant at the Institute of Journalism and Public Relations at the FH JOANNEUM University of Applied Sciences in Graz, Austria. Her interests lie in data journalism and digital media.

Eva Goldgruber is a research fellow and lecturer at the Department of Media and Design at the FH JOANNEUM University of Applied Sciences in Graz, Austria. Her areas of interest are education, technology, communication, and their intersections.