

# Status quo and challenges of data journalism

A literature review on specific domain problems

Authors:

Julian Ausserhofer

Eva Goldgruber

Robert Gutounig

Sarah Matiasek

Thomas Wolkinger

Date: 3.5.2017

Project title: Visual Analytics in Data-driven Journalism

This work was supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) under the ICT of the future program via the VALiD project no. 845598.

This document was developed in the context of the research project “Visual Analytics in Data-driven Journalism (VALiD)”(<http://www.validproject.at>), supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) under the ICT of the future program via the VALiD project no. 845598; RTI initiative: „ICT of the future“.

### Project Synopsis:

Data journalism extracts complex information from a vast amount of data and presents it visually. However, there is often a lack of suitable analysis methods and the appropriate technology for this difficult task. The research project "VALiD - Visual Analytics in Data-Driven Journalism" develops new techniques to support journalists in this task. These techniques can be used, for example, to analyse data from parliamentary speeches or the media transparency database which then can be presented visually. The goal of the VALiD project is to significantly improve those visual analytics technologies that support the investigation of complex and heterogeneous data in data journalism. Three main hurdles exist and are tackled by the project: usable systems, the appropriate technology and the missing knowledge and workflow in the work routine.



# Contents

## [1 Executive Summary](#)

## [2 Introduction](#)

### [2.1 The Datafication of Journalism](#)

### [2.2 Research questions](#)

### [2.3 Method: Systematic Literature Review](#)

#### [2.3.1 Literature retrieval and selection](#)

## [3 Analysis of Research Literature on Data Journalism](#)

### [3.1 Defining Data Journalism](#)

### [3.2 Job Titles and Job Descriptions](#)

### [3.3 Influential Works for Data Journalism Research](#)

### [3.4 Research designs](#)

### [3.5 Characteristics of data journalistic pieces](#)

### [3.6 News Production Workflows in Data Journalism](#)

#### [3.6.1 Integration of data journalists into the newsroom](#)

#### [3.6.2 Data Journalism Workflow Models](#)

##### [3.4.2.1 The Inverted Pyramid Model by Bradshaw](#)

##### [3.4.2.2 Data Journalism Workflows at The Guardian](#)

##### [3.4.2.3 Collecting the data and finding a topic](#)

###### [3.4.2.3.1 Different countries, different practices](#)

##### [3.4.2.4 Analyzing](#)

##### [3.4.2.5 Building and Visualizing](#)

##### [3.4.2.6 Collaboration](#)

##### [3.4.2.7 Publishing the Product](#)

### [3.7 Backgrounds of Data Journalists](#)

#### [3.7.1 Educational Background](#)

#### [3.7.2 „Computational Journalism has no textbooks yet“ - Educational initiatives](#)

#### [3.7.3 Necessary skills for data journalism](#)

### [3.8 Enabling and Hindering Factors of Data-driven Journalism](#)

#### [3.8.1 Accessibility](#)

#### [3.8.2 Collaboration and Position in Newsroom](#)

#### [3.8.3 Tools](#)

#### [3.8.4 Supportive Infrastructure](#)

#### [3.8.5 Lack of Resources: Time and Money](#)

#### [3.8.6 Business Models & Revenue](#)

### [3.9 Gender aspects in Data Journalism](#)

### [3.10 Research gaps](#)

## [4 Conclusions & Outlook](#)

## [5 References](#)

## [6 Appendices](#)

[6.1 List of definitions for data journalism](#)

[6.2 Literature Corpus](#)

# 1 Executive Summary

This report summarizes findings from academic research on data journalism. The increased use of data in journalism can be seen in connection with the current datafication of social spheres in general and the transformation of journalism in particular. By conducting a Systematic Literature Review we tried to answer how journalistic workflows and routines have been affected due to the broad introduction of data in the newsroom, how this impacted journalistic practices and skill requirements for journalists, what user needs, tasks and goals in data journalism can be identified and finally what role gender plays in current practices of data journalism.

Reviewing the existing literature it can be noticed that already agreeing on a common definition of data journalism seems an almost impossible task. Therefore a wide range of definitions were contrasted in order to frame the practice of data journalism and also to identify the most influential works on this topic. We analyzed a range of applied research designs in order to find signposts for future research directions. Data journalism pieces are united by a set of characteristics which we tried to single out, such as interactivity, simplicity etc.

The main part of the report is devoted to the analysis of news production workflows in data journalism. From collecting the data, analyzing, building and visualizing, collaboration patterns to publishing the final product commonly ascribed characteristics were identified.

In a further part we described backgrounds of data journalists in terms of education, educational institutions for data journalism and desired skills (e.g. coding, data-related skills etc.) as well as common enabling and hindering factors (e.g. accessibility to government data, lack of resources etc.). Finally aspects of gender in data journalism were focused and possible research gaps identified (e.g. cross-national investigations of data journalism).

## 2 Introduction

### 2.1 The Datafication of Journalism

In 2013 journalism was already in a worrying state: due to technological development newsrooms were closing all over the place: “The disruption that the Internet has posed to the traditional business models of newspapers has been well-documented over the past decade. More than 166 U.S. newspapers of an estimated 1,382 in total have stopped putting out a print edition or closed down altogether since 2008, resulting in more than 40,000 job losses or buyouts in the newspaper industry since 2007” (24<sup>1</sup>). Times have certainly changed and newspapers do not hold a monopoly any more. Also in other countries newspapers are continuously struggling with new technology and are facing a paradigm shift. Until now, no profitable business model for online media has been found even though the issue is discussed continuously on panels and conferences (24).

However, not only newspapers had and have to adapt to the technological changes. Also governments, research institutions, non-governmental organizations and businesses went online and started to make data publicly available (20). Among the forerunners of open government and open data certainly are the United States and the United Kingdom. Therefore, it is no wonder that the most prominent examples of successful data journalism can be found in these countries: the Guardian and the New York Times (32).

The aforementioned development and the shift towards online media opened another door for journalism. As Mirko Lorenz said: “The big worldwide market that is currently opening up is all about transformation of publicly available data into something that we can process: making data visible and making it human. We want to be able to relate to the big numbers we hear every day in the news - what the millions and billions mean for each of us” (24).

---

<sup>1</sup> Numbers in brackets refer to items from the research literature corpus which is listed in the Appendices section.

## 2.2 Research Questions & Goals<sup>2</sup>

In this report we reflect on the published research on data journalism in a structured way and provide a signpost for future research paths. In particular we want to deal with the following questions:

1. In what way have journalistic workflows and routines changed due to the broad introduction of data in the newsroom?
2. How has this impacted journalistic practices and the skill requirements for journalists?
3. What are identified users' needs and challenges in data journalism?
4. What role do gender aspects play in current practices of data journalism?

Each of these research questions is addressed below and lays the groundwork for a major aim of this report: Providing propositions for future research on data journalism.

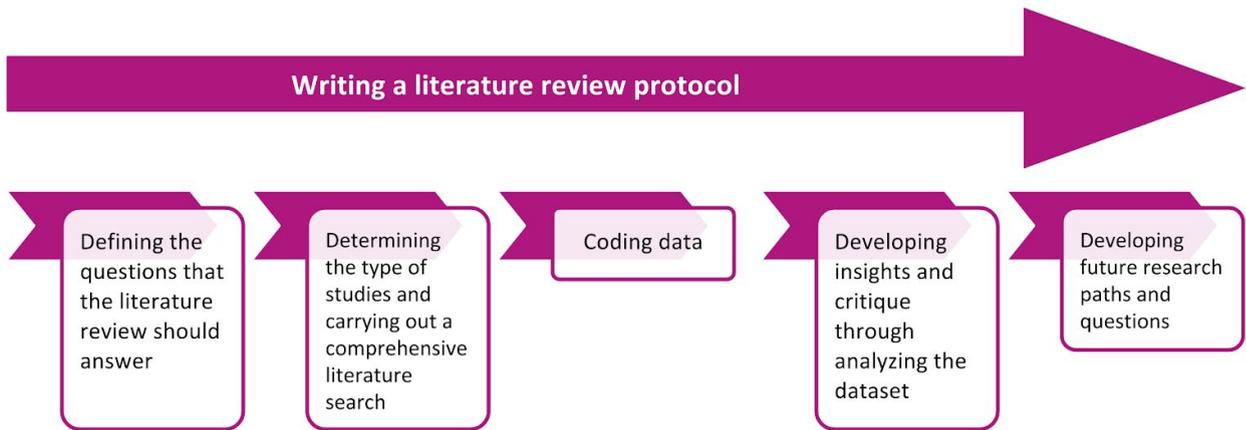
## 2.3 Method: Systematic Literature Review

Concerning the methodological frame we were following the concept of a structured literature review (SLR) which, in the literature review continuum, starting from a rapid review represents the most rigid form of a literature review to date (Massaro, Dumay, & Guthrie, 2016). The goal of a SLR is "to develop insights, critical reflections, future research paths and research questions" (Massaro, Dumay & Guthrie, 2016) and it should "contribute to developing research paths and questions by providing a foundation on which to build on prior discoveries." (Massaro, Dumay & Guthrie, 2016) On the one hand, they provide the basis and the justification for new research, on the other hand, for more mature areas of research, they "provide the background to develop research synthesis." (Massaro, Dumay & Guthrie, 2016) Systematic reviews (as well as SLRs) adopt "a replicable, scientific and transparent process [...] that aims to minimize bias [...]" (Tranfield, Denyer & Smart, 2003). SLRs are based on a "positivist, quantitative, form-oriented content analysis method for reviewing literature", however, they also strongly use hermeneutic and interpretative methods, especially when developing insight and critique (Massaro, Dumay & Guthrie, 2016). The purpose of critique in this context "is to counteract the dominance of taken-for-granted goals, ideas, ideologies and discourses [...]" (Alvesson & Deetz, 2000).

---

<sup>2</sup> This report served as a basis for results published in Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (forthcoming). The Datafication of Data Journalism Scholarship: Focal Points, Methods and Research Propositions for the Investigation of Data-intensive Newswork [Accepted Manuscript]. *Journalism*.

Figure 1  
Undertaking a systematic literature review



The undertaken steps are illustrated in fig. 1: First we defined the research questions the literature review should answer. Then we were following a rigid and reproducible path in selecting the relevant literature which we subsequently analyzed and coded. In order to gain more insights from the literature we enriched the content analysis with quantitative analysis, for example citation metrics. This data was retrieved from Google Scholar which represents “a valuable source for assessing impact when conducting an SLR and can help provide valuable insights and critique of specific issues in a research field.” (Massaro, Dumay & Guthrie, 2016)

### 2.3.1 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; Lecheler and Kruike-meier, 2016; Massaro et al., 2015; Neuberger et al., 2007). For a special version of this report, a research paper, we decided to only include empirically grounded papers, irrespective of whether the investigation focused on quantitative, qualitative or mixed-methods research. For this report we also included works with a predominantly conceptual focus. Our focus was on publications from the social sciences, but we also included papers from other disciplines (e.g. computer science). In terms of publication type, we considered journal articles, book chapters, conference papers, industry reports and PhD theses. In some cases Bachelor's and Master's theses, press reports, and blog posts, as well as popular textbooks.

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 papers. This led us to the following list of synonyms for data journalism:

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

Synonym	
algorithmic journalism data-driven reporting computational journalism database journalism computer-assisted reporting	datajournalism data journalism Datenjournalismus data-driven journalism quantitative journalism

*Note.* While the German term (‘Datenjournalismus’) was included, the most widely used French and Spanish expressions (‘journalisme de données’, ‘jornalismo de dados’) had to be excluded due to insufficient language proficiency. Other related terms, more general concepts and abbreviations such as ‘accountability journalism’, ‘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.

Based on the synonymic search terms from table 1, we undertook a literature search in 15 scientific databases, obtaining 772 records in total. The search terms were combined with the Boolean ‘OR’ operator.

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

Database	Number of records		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

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

The records were imported into Zotero (Roy Rosenzweig Center for History and New Media, 2015) and assessed independently by two researchers considering title, abstract and keywords, and selected according to the above-mentioned formal criteria and research focus. 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.

Two further measures were taken in order to select the literature for the corpus. Firstly, we asked three domain experts, all researchers in the field of data journalism, to add relevant work based on their expertise. Secondly, we checked the references of our selection, computer-supported and manually: We scraped the references of all selected papers 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 papers. Lastly, the references of the selected works published in 2015 and 2016 were checked by hand. This process resulted in a corpus of 40 publications.

The final corpus of research publications allowed us to perform different types of quantitative and qualitative analyses. Since research publications are rich in both structure and content, we decided to analyze both dimensions of our corpus.

We conducted a software-assisted qualitative content analysis (Kaefer et al., 2015; Mayring, 2000; Schreier, 2012), starting with the development of a code book. While the code book’s main categories were defined deductively by our research interest and research questions, we derived the subcategories out of the material manually. Such an inductive as well as deductive 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 papers, continuously discussing and modifying the category system. They then compared the coding and discussed differences

until an agreement on the codebook was reached. We used the qualitative data analysis software Nvivo (QSR International, 2015) for the coding.

Finally, we supplemented the manual coding with an automated content analysis which used a python implementation of the TF-IDF algorithm to extract the central terms of each paper. TF-IDF, or term frequency-inverse document frequency, is a statistical measure that evaluates the importance of a word within an individual document, as well as the corpus as a whole. During the course of the project a Literature Explorer was developed which helps to find out certain characteristics of the literature corpus on data journalism. For further perspectives on the corpus, please also visit <http://literature.validproject.at>

## 3 Analysis of Research Literature on Data Journalism

### 3.1 Defining Data Journalism

In a newly emerging field it is common that there are no fixed definitions yet and people have different concepts in mind when using the same term. This phenomenon can also be detected in the field of data journalism. Therefore, we would like to clarify what we mean by data journalism. Apart from synonyms extracted from literature's keyword section (table 1), the practices which can be in some way or the other related to data journalism run also under following labels: data journalism, data-driven journalism, data data-driven investigative journalism, data-driven interactive storytelling, precision journalism, computational journalism, computer assisted reporting, database journalism, structured journalism, journalism as programming, programming as journalism, news applications etc (14, 18, 24, 31, 33).

Some of the definitions give a focus on the change in the process of newswork (Appelgren and Nygren, 2014; Bounegru, 2012; Davenport et al., 2000; Diakopoulos, 2011; Felle, 2016; Flew et al., 2010; Karlsen and Stavelin, 2014; Lorenz, 2010; Parasie and Dagiral, 2013; Tandoc Jr. 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, e.g. maps or diagrams (Baack, 2013; Hullmann 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, 2012; Ausserhofer, 2015; Coddington, 2015; Cohen, Hamilton, et al., 2011; 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 is preferred: interactive visualizations. To capture this diversity, this report emphasizes the process and the product equally.

Prototypical examples of data journalism include the Guardian's ‘Afghan’<sup>3</sup> and ‘Iraq War Logs’<sup>4</sup> (mentioned in Baack, 2013; Knight, 2015); ‘Chicago Crimes’, 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 using a variety of terms (see table 1). Many of them are connected with different communities, histories, epistemologies and visions of the public (Bounegru, 2012; Coddington, 2015). However, what is important to this report 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 (an umbrella term we want to suggest here) 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.

For a list of how authors define the most commonly used terms see Appendix.

We also want to mention that the literature suggests two distinct forms of data journalism:

- Investigative Data Journalism (IDJ): Journalists have plenty of time for their work and spent months up to years on a project.
- General Data Journalism (GDJ): This happens in the daily routine at a newspaper and journalist only have a few hours time (19).

---

<sup>3</sup> <https://www.theguardian.com/world/the-war-logs>

<sup>4</sup> <https://www.theguardian.com/world/iraq-war-logs>

## 3.2 Job Titles and Job Descriptions

Data journalists figure under many different names and descriptions. Among the labellings used in the research literature the following can be found (04, 06, 11, 12, 13, 17, 18, 19, 24, 27, 30, 31, 33, 39, 40) :

- Journo-coders
- Journo-devs
- Jourveloper
- Journalist-programmer
- Programmer-journalists
- Editorial programmer
- Progo-journalist
- Coder-journalist
- Hacker-Journalist
- Computational journalist
- Computationalist
- Computer reporter
- Computer-assisted reporters
- Computer assisted reporting specialist
- Data reporter
- Data journalist (“Datenjournalist”)
- Data person
- Data editor
- Database editor
- Development editor
- Data visualisation expert
- Interactive news editor
- Interactives journalist
- Infographic design editor
- Information graphic designers
- Graphic editors
- Art and design directors
- Editor of news application
- News application developer

### 3.3 Influential Works for Data Journalism Research

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 towards an academic audience. Phil Meyer's (Meyer, 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 paper 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 papers in our corpus. Other often-cited and influential works on data journalism can be seen as part of research that has 'focused on the technological promises of computing on journalism' (Gynnild, 2014: 714), or publications that explore the potential of the practice (Cohen, Hamilton, et al., 2011; Cohen, Li, et al., 2011; Flew et al., 2012; Hamilton and Turner, 2009). Table 3 lists the ten most-cited publications in our corpus.

**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, Hamilton, et al. (2011)	8
Cox (2000)	7
Cohen, Li, et al. (2011)	7

Lewis and Usher (2013)	6
------------------------	---

### 3.4 Research designs

The analysis of research designs is summarized below in table 5. It revealed a certain scarcity of quantitative research designs and digital methods, in contrast to qualitative exploratory ones. Less than one in 10 studies we analyzed collected their data using surveys. 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 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 one case over 100 (Howard, 2014). In total, 25 publications used interviews as a method of investigation. 7 of those questioned 5 to 9 individuals for their research, 6 interviewed between 10 and 19, and 6 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 many publications employed mixed-method approaches. 8 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 artefacts.

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 a paper by Lugo-Ocando and Brandão (2015) that analyzed news items produced by journalists in the UK, and one publication, which analyzed articles within the Guardian's Datablog with respect to news values, sources and topics (Tandoc Jr. 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 discussed 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 data blog – and online

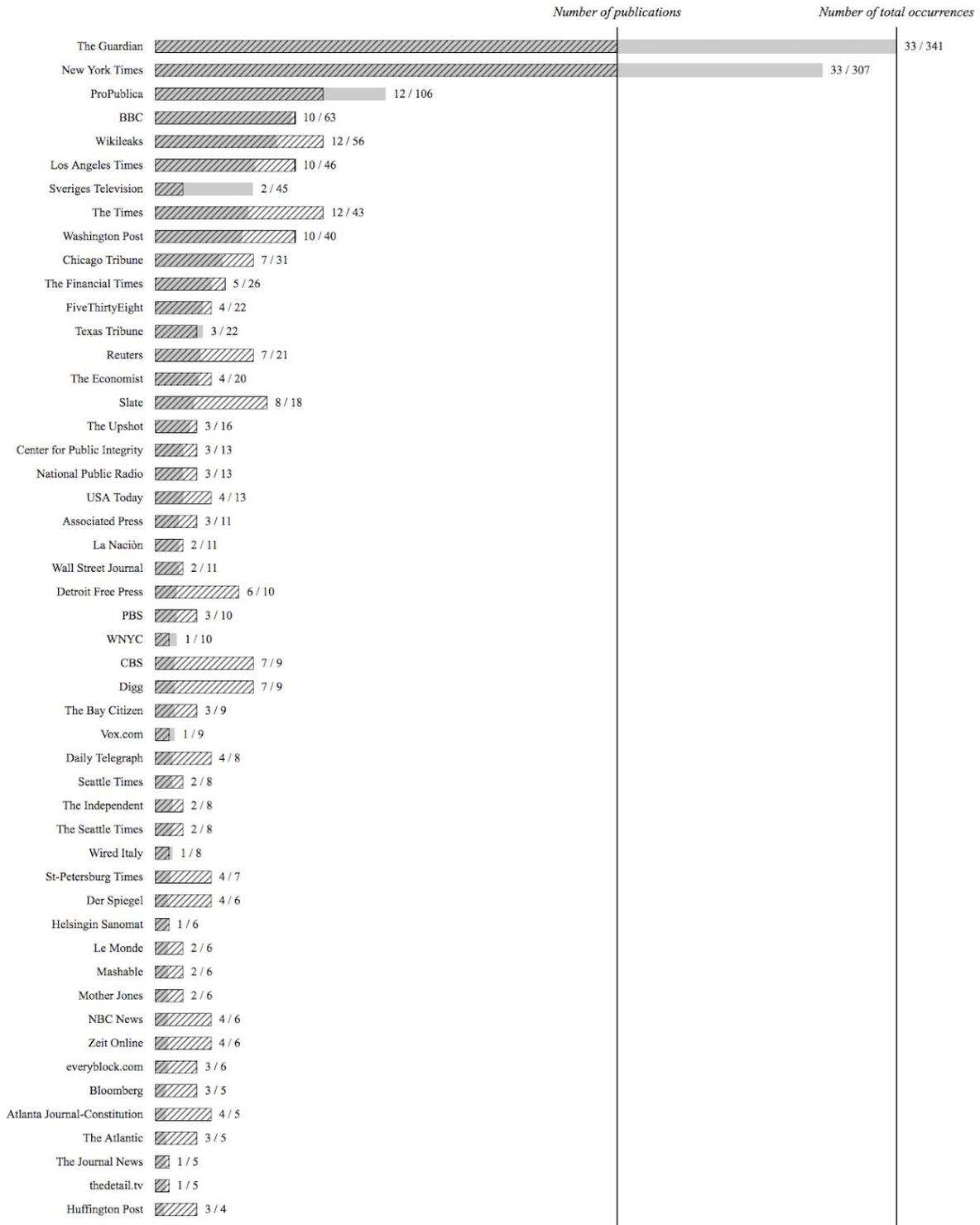
news accounts, list serves and journalism blogs to grasp the discourse around data journalism.

The method of survey was used less frequently, though (Appelgren and Nygren, 2012, 2014) 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 eight hours observing the BBC News Online Specials team.

Table 5. Research methods of publications on data journalism

**Figure 2.** Occurrence of media organizations.



Which media organizations did researchers associate with data journalism and investigate? The list in figure 2 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 smaller news organizations (e.g. Mashable, Mother Jones etc.) that were only rarely

mentioned but might suitable for future research could also be identified. In the German speaking countries media organizations such as Zeit Online, Berliner Morgenpost, Neue Zürcher Zeitung, Der Standard and Kurier have also issued interactive infographics which can be seen as examples of data-intensive newswork (Zwinger & Zeiler 2016).

### 3.5 Characteristics of data journalistic pieces

Talking about the final products of data intensive newswork researchers have also attempted to identify common characteristics of the results. A common feature seems to be interactivity which can be classified according to its degree (low, medium or high) or by course of action (linear, nonlinear or linear-nonlinear) (Weber & Wenzel 2016). People working in this field described the “ideal of ‘drillable’ interfaces to data where both the overall and the detailed view are represented” (03). They also emphasize that the end product must be designed for the lay user not the data expert and is often designed in a self-exploratory way (03, 33). Consequently many journalists focus on design elements supporting a linear narrative rather than free exploration (03), while in general data journalistic pieces also allow for non-linear navigation ways (33). Often there are only one or two input fields available for the user (33). There seems to be “a preference for simplicity and clarity” (03) when narrating data stories.

While representations of data are at the core of data intensive newswork, they are often counterbalanced by the use of quotes from individual witnesses or experts who also sometimes take over the role of explaining the figures (06). Prototypical data stories mentioned by people working in the field often have personal characteristics “connecting data with the lives of their audience.” (39)

There are projects which provide a snapshot of a data situation, some are ongoing and provide continuous update of data (08). However, ongoing projects often face problems of maintenance or uncertainty regarding who is responsible for the project (08).

### 3.6 News Production Workflows in Data Journalism

#### 3.6.1 Integration of data journalists into the newsroom

The integration of data journalists or programmers into the newsroom is a crucial issue. Those media organizations which already focus on data-intensive newswork handle the structure of teams differently, depending on their size and resources. Newspapers that have

enough financial resources hire experts and establish special departments to meet the needs.

Research shows that there seem to be three different models (19):

- traditional data desk
- flexible data projects model
- entrepreneur model

The names of these departments differ greatly but what they do have in common is the understanding that it is important for efficient and effective work to have programmers and journalists work together on the projects. Today most large newsrooms contain information-visualization departments. The respective teams are labelled as “news applications team” (06), “interactive news team” (19) “interactive news technology department” (03), “data desk” (11, 16). All of these namings highlight certain aspects of data intensive newswork. The New York Times has its own Interactive News Technology Department while at the BBC, the data journalism unit is called “News Online Specials Team”. At The Times the Visual Journalism Unit is responsible for interactives. The Guardian even employs two interactive teams, one solely for the US market. In smaller newsrooms, however, it is still common for journalists to produce their own visualizations (31, 09).

“In German and Swiss editorial departments it is still the journalist or the editor who is responsible for the content or the story of the interactive graphic; the designer only delivers his visual part, the same goes for the programmer. An interviewed designer says: ‘Mostly the journalist does the research, and we care for the visual things like maps or charts.’ In contrast, the interviewees of the New York Times stated: ‘We are journalists’. In this statement, we recognize a paradigm shift that has occurred in the New York Times newsroom. What is new is that even programmers and designers see themselves as journalists; they belong to the journalistic team of the newsroom and define their task as a journalistic task” (12).

### 3.6.2 Data Journalism Workflow Models

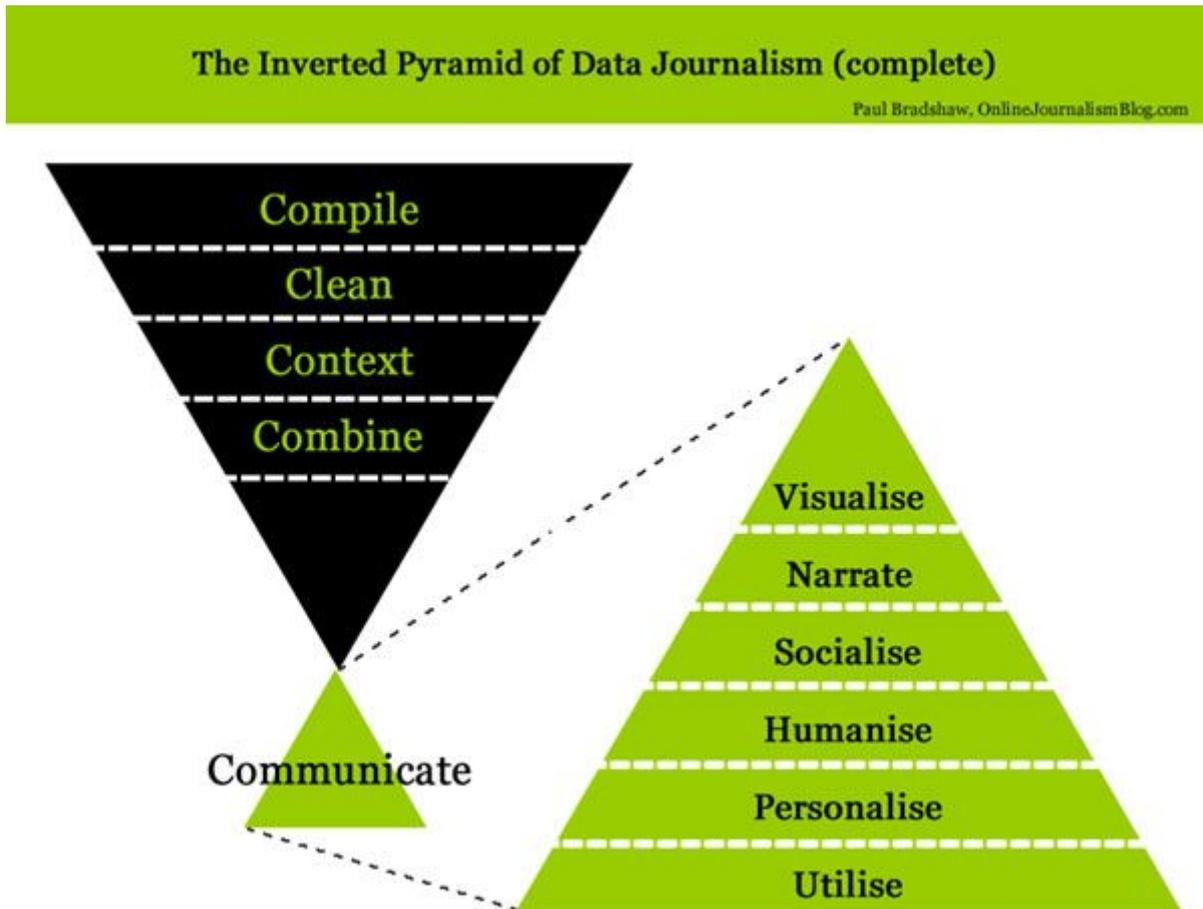
In our literature corpus different steps for the production of a data journalism pieces are suggested or observed. Even though there is no fixed workflow in data journalism, there are

certain steps practitioners and scholars seem to agree on. “In summary, the workflow can be roughly divided into following stages (27):

1. Identifying the gist for the story and the potential role for the data in the story.
2. Identifying and obtaining the right data sets to respond to journalists’ questions.
3. Modifying the data to make them ready for analysis — e.g., correcting errors in the data sets.
4. Analysing the data with the right tools, and mashing the data with other data sets if relevant.
5. Producing the story: text, visualisations, interactive elements.
6. Publishing the data sets that were used in the analysis.
7. Inviting readers to participate by reusing the data, commenting on and sharing the story through applications in social media, and submitting more content through applications like Flickr, Facebook and Twitter.”

#### 3.4.2.1 The Inverted Pyramid Model by Bradshaw

Paul Bradshaw, online journalist and blogger, also developed a model for data journalism based on the idea of the inverted pyramid.

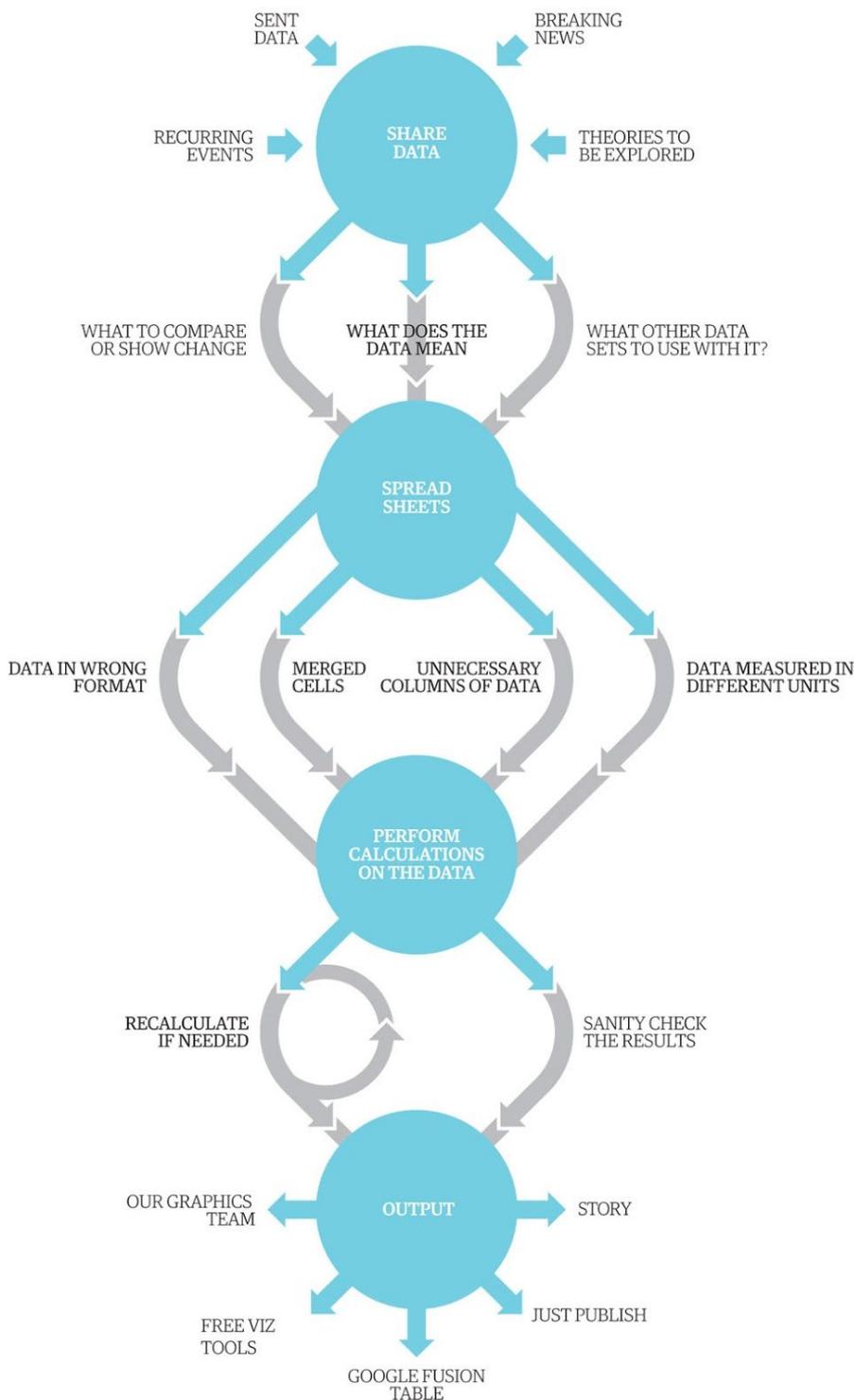


Source: Bradshaw, Paul (2011).

Bradshaw divides the process into five steps. The first step is compiling the data which is followed by cleaning. In the next step the data has to be looked at critically: “who gathered it, when, and for what purpose? How was it gathered? (The methodology). What exactly do they mean by that?” In addition, it is important to familiarize with the data, “understand jargon, such as codes that represent categories, classifications or locations, and specialists terminology” (Bradshaw, 2011). This step might, according to Bradshaw, lead back to compiling new data or to the next step, which is combining. “Good stories can be found in a single dataset, but often you will need to combine two together” (Bradshaw, 2011). The last step is communication the data. For this step, Bradshaw suggests another pyramid, as can be seen in the picture.

#### 3.4.2.2 Data Journalism Workflows at The Guardian

The editors of the Guardian’s Datablog, “an extensive data journalism Web section” (32) established in 2009, have developed their own workflow which they illustrate in the following graph:



Source: Rogers, S., & McCormick, M. (2011, April 7). Data Journalism Broken down: What We Do to the Data Before You See It. The Guardian. Retrieved from <http://www.theguardian.com/news/datablog/2011/apr/07/data-journalism-workflow>

At the Guardian a story usually starts with the data which can stem from various sources. Then the editors decide how they want to treat the data, and what they want to do with it. Sometimes they combine several data sets, sometimes they focus on changes over time.

Before calculations can be performed, the data has to be checked, faulty data needs to be cleaned. As a last step, the editors decide what output they want to achieve and handle the data accordingly (Rogers & McCormick, 2011).

The following workflow steps are derived from our literature corpus:

#### 3.4.2.3 Collecting the data and finding a topic

Before the very first step in the data journalism workflow - collecting the data - can be taken a topic has to be found. Several topics haven proven to be especially interesting for data journalism as they provide journalists with enormous amounts of analyzable data and pose interesting questions. Research shows that topics especially interesting for data journalism include (06, 09, 11):

- Demographics
- Crime
- Politics
- (Private) economy/finances
- Health care
- School/education
- Environment
- Business and economy
- Working conditions
- Urban services and transportation
- Culture
- Sports

Science is relatively underrepresented as a topic in data stories, but data elements are a substantial element of science stories (10, 15).

Generally, there seem to be two primary ways to start a data journalism story. "In the first option, a data set provides more information for a story topic that was already discovered by the newsroom. In the second, a data set serves as a starting point for the whole story" (27). A common starting point is data released by statistical institutes (18). Projects could also be event-based (39). At the New York Times, for example, most projects originate in stories

where data is available as well. They also do projects where they “recognize a need”, where they produce “more of a tool-based rather than a story-based project” (33).

In other words, three approaches can be identified (39):

“1. Reporting-team-driven data: Sometimes a team leader will say, what kind of data do you have to bring to bear on question?

2. Issue/news-driven data: Sometimes you take a running look at data, like campaign-finance data as it flows in over the next six months or on health care industry money.

3. Data-team-driven data: Data team approaches a topic because it’s a hot issue” (39).

“Because developers and reporters often have specialized skill sets, it’s important to bring them together to brainstorm story ideas. This increases the likelihood of identifying unique and significant angles for data-driven stories.”

When an interesting topic has been found, collecting the data is the very first step in the workflow. There are several ways to maintain necessary data. This process may vary from newspaper to newspaper as well as from country to country. First and foremost, governments and organizations prove to be a major source for data. However, “news organisations use data sets which they have collected from several sources and then compiled themselves” (27).

Common sources/origins of data in journalism include:

- Official sources: government (06, 07, 08, 10, 13, 19, 27, 28, 29, 39, 40)
- Freedom of Information Request/Act (FOIA)/Access to Information Request/Act (03, 07, 10, 18, 27, 37, 40)
- Journalist’s research, compiling surveys, polls (03, 06, 10, 13, 27, 29, 39, 32, 40)
- Crowdsourcing: collecting data from the audience or engage them in a topic (03, 08, 13, 15, 17, 32)
- Leaks (10, 13, 27, 19, 30)
- Technical methods: scraping, data mining (13, 14, 18, 27, 29)
- Non-commercial organisations: universities, research institutions, NGOs, Wikileaks (10, 27, 29)
- Press releases (10, 13, 34)

- Big data (19, 32)
- Electronic public records (03, 36)
- Breaking news stories (39)
- Corporate entities (10)
- Pan-national organisations (such as the various agencies of the United nations) (10)

#### 3.4.2.3.1 Different countries, different practices

**Norway:** According to Norwegian journalists the access to data is satisfying. “In Norway access to public data is regulated by freedom of information legislation that favours transparency and public inspection.” However, “while data access offers few problems” the commercialization of public data in Norway seems to be an issue. According to the literature, “some governmental bodies are allowed to charge money for data, and put barriers around the data in ways the current legislation does not (arguably) address.” In addition, data journalists in Norway make use of crowdsourcing to gain data (03). “In Norwegian newsrooms, teams consisting of people with different skills create this form of journalism. They still often depend on central programmer-journalists, people who can program but indicate their profession as being journalists or data journalists” (31).

**USA:** While journalists in the US (Chicago) “analyzed databases that they build themselves from data they had collected” (06). themselves back in the day, they started analyzing “existing databases released by the government or built from publically available data” (06). Publicly available data released by authorities is a main source. “As governments release data in computerized form, the computer-assisted reporter seldom designs a new database (only for 7.3% of stories). Moreover, most published articles are based on data released by public officials (60.9%). So most stories are driven by the releasing of data, and the reporter identifies paths in those data, based on statistics” (06). In general, news organizations in the US seem to gain huge amounts of data from governmental agencies either drawing conclusions from already published data or via filing a Freedom of Information Act (FOIA) (06, 07). Those with good infrastructure, technical and skill wise, may also use technical methods like scraping to get data (27).

**Sweden:** In Swedish newsrooms crowdsourcing is an important way of collecting data for data journalism activities. Journalists believe that “most data submitted by the audience are of good quality and trustworthy. [...] The editor refers to such data as a form of ‘wisdom of the crowd’” (08).

**UK:** In the UK several sources of data are used. Amongst them are governmental sources, corporate entities as well as research institutes, non profit as well as academic institutions, “Pan-national organisations (such as the various agencies of the United Nations), polls and self-generated (i.e., data gathered by the news organisation itself)” (10). Research shows that data journalists in the UK “relied heavily on institutional sources, especially government agencies. In addition, there is evidence of the rise of data-based press releases: a substantial proportion of the stories showed evidence of a body of data being released wholesale by research institutes and other vested bodies. Particularly in social issues and health, most of the data presented were acquired in this way” (10). However, also self-generated data, leaked data and freedom of information request are common (10). For example, “the Guardian successfully crowd-sourced on a number of major data stories, the most famous being its readers sifting through thousands of documents on expenses claimed by their Members of Parliament (Rogers, 2009)” (15). Readers are asked to help “analyze data sets, provide feedback or additional data, formulate research questions, and submit applications and visualizations that they have created from accessible data in the Data Store” (18). (10, 15, 18)

**Germany:** In Germany reported data collection methods include leaked information, journalist's own research, crowdsourcing, scraping techniques or through evaluation of already published data from open government or press institutions (13). Often data is also stored in case it proves to be useful at a later point (14).

**Canada:** In Canada information requests are also common: “However, this procedure has constraints that greatly limit its application. At the federal level, the Access to Information Act that regulates this right stipulates that fees may be charged for document search, processing and reproduction” (40). A study published in 2016 shows a very small number of projects based on self-build databases. Public data is the main source for journalists. Some are able to use their technological knowledge to “program algorithms and, more specifically, scripts, to collect data from a government site or image file” (40). “One considerable barrier to the production of quality data journalism in Quebec is the lack of quality data offered by the provincial and federal government” (38).

#### 3.4.2.4 Analyzing

In data journalism collecting the data is obviously not enough. While “journalists ask questions; data journalists ask the data” (09). Before any journalistic enquiry data sets have to be filtered, selected and errors have to be dismissed (03). Cleaning the data can be very exhaustive and repetitive. As Jonathan Stray, an instructor at Columbia Journalism School, points out: this “is often the most time consuming part of the data journalism process” (24). Often datasets are messy or they are delivered to the journalists in non-machine-readable formats. In many cases PDF documents hold valuable information but due to the fact that they would have to be converted manually into a machine-readable format, they often cannot be used (32). As this step would take up lot of time - which is generally a rare resource in the newsroom - the story that could have been told is lost.

In an explorative phase at the beginning journalists discover what kind of data they are working with and also in what condition the data is. They further figure out if it is worth it to work with the data (13). Generally, patterns within the data have to be identified. The main goal is to find a story within the data. At the beginning the data is roughly cleaned, hypotheses and a storyline are formed (14). However, there is no clear way in this process, it may vary from story to story (13). Sometimes “the combination of heterogeneous data sets may generate useful knowledge” (06). “In the vast majority of cases, the data is analysed with a focus on comparing values (e.g., to show differences between men and women or neighbourhoods) and almost half of the pieces show changes over time (e.g. “Climate Change: How Hot Will It Get in My Lifetime?”) (29).

Some data sets are too vast to base a single story on them. Other data sets need to be combined with others to tell an interesting story. “There are general rules to follow when putting together a database and making sure that the quality of the data is good. The fundamental activity is to normalize the database using SQL (Codd 1970). When doing this, errors in the data are often discovered.” (03). The ultimate goal is to find a journalistically valuable story within the data (13). Other methods include the use of statistical software like Excel, Access, SPSS or R and collaboration with users called crowdsourcing (14). Crowdsourcing can provide interesting insights. At the Guardian, for example, “readers are encouraged to help analyze data sets, provide feedback or additional data, formulate research questions, and submit applications and visualizations that they have created from accessible data in the Data Store” (18).

Messy and inaccurate data can be problematic. Sometimes the same entities are labelled differently within the same data set (07). “No information is ever clean. No data is ever perfect. I’m willing to accept that. But given that limitation, it’s really important and really necessary and really hard to figure out what we can responsibly say with this information. (Armendariz, interview, September 12, 2012)”

#### 3.4.2.5 Building and Visualizing

A key element is the visualization of the data: “The use of images not only helps in attracting readers’ attention but also affects reading experience (Ensenberger 2011).” (32) When it comes to visualizations, different strategies are used depending on the complexity of the story. Numerous helpful and free tools exist today that can help journalists produce appealing graphics and visualizations by themselves without the help of professionals like designers or programmers. However, sometimes the skills of an expert are needed to meet the needs of the project (01, 27).

#### **Tools used for visualizations (01, 11, 22):**

- Data visualization programs like Caspio and Tableau
- Mapping programs like ArcView
- TimelineSetter: “an interactive HTML timeline tool that turns ‘ugly spreadsheets into pretty timelines’ (Phelps, 2012)” (22)
- IBM’s ManyEyes: allows easy upload and visualization for free
- Datawrapper: for building simple diagrams and graphics
- Google Fusion Tables
- MySQL, Access, Excel
- CartoDB

The design of an infographic is very important as the interest of the readers depends on it. If there are too many buttons readers might be overstrained and as a consequence they do not pay attention to the graphic and its story (03).

Most commonly, data stories are visualized with one or more of the following (03, 06, 10, 29, 33, 39):

- Pictures

- (Simple static) charts
- Maps (static or interactive)
- Tables
- Infographics
- Timelines
- Animated visualisations (rarer)
- Lists (of numbers)
- Text (as numbers are discussed within the text, but not represented in other ways)
- Sound and video
- Online presentations

Several factors have to be taken into account when designing an interactive infographic. Psychological research on visual salience found that exceptions from the norm attract viewers attention: “The strength of this attraction is modulated by multiple factors, including the scene itself (e.g., a brightly colored object is less salient when surrounded by other brightly colored objects) and by the viewer’s task (e.g., expectations and top-down search can affect what is perceived as most salient).” (35) Also cultural factors have to be taken into account. Often, the starting point of a visualization is a storyboard, which illustrates “the graphic through a series of sequentially arranged images” (12). This also helps journalists and developers in their communication with each other as well as the production process (12).

Other common routines are wireframes, sketches and drafts drawn on a computer. Those shall help to integrate the view of the end user into the concept of the visualization (14). As with storyboards, wireframes support the coordination between journalists and developers. “Wireframing typically occurs in a graphics program like Illustrator or other design program, in order to immediately engage the visual aspects of a project. The wireframe gets a full review, by editors, others on the team and the associated desks involved in the project. Once the wireframe is complete, the site is built out, and depending on size, it is staged on a Web server and tested.” (33) In addition, physical and electronic project coordination tools, like pinboards, which give information about the status quo of the project as well as shared chats or Skype, Wikis and mailing lists can help. Division of work within a data journalism project is facilitated by supportive infrastructure (14).

An important step in building a visualization for a story is the decision of whether to build an experimental infographic or to take a “prescriptive narrative approach”, which seems to be preferred by readers (17, also 03). However, The Chicago Tribune like many other programmer-journalists for example “also put the emphasis on giving out the data - through friendly applications - so that the readers can do their own research on how an issue affects their personal situation” (06). In addition, the ability of readers to read such information graphics has to be taken into account (06).

Before visualizations are published, internal tests and feedback loops are performed. A further important step is the adaption of visualizations for different platforms, operation softwares and browsers and also mobile use (14).

Even though a lot is possible in data journalism today, there still can be found many cases where “already-assembled datasets” are simply visualized with automated programs without further analysis (40).

#### 3.4.2.6 Collaboration

Collaboration plays an important role in data-driven journalism. Often work is carried out in teams (e.g. 03). As previously mentioned, some media organisations have already established interactive departments where people with different backgrounds work together (03, 20). However, in Norway this is not the case yet. Journalists here are reported to work “alone and in small teams, figuring out how to best collaborate as they go” (03). A study shows that “data journalism, more often than not, is a collaborative effort” (29).

Collaboration can take on different forms within the process. “In German and Swiss editorial departments it is still the journalist or the editor who is responsible for the content or the story of the interactive graphic; the designer only delivers his visual part, the same goes for the programmer.” (12) In contrast, at the NYT “even programmers and designers see themselves as journalists; they belong to the journalistic team of the newsroom and define their task as a journalistic task” (12). However, it is reported that in German and Swiss media companies tendencies to overcome this gap exist. Along those lines, special teams consisting of people with different skills are built to work on special projects (12). Research shows, that “a team of at least three experts has proven itself in practice: a programmer (e.g. flasher), a designer, and an author” (12). Complementary backgrounds favourable for a multimedia project seem to be key: “The teams generally consist of a combination of skills in

journalism, Web development, data analysis, visualisation and statistics. Typically, the leader of a data team is a journalist with data-analysis skills. The rest of the team are alike - either journalists or nonjournalists with skills in statistical analysis, data processing, visualisations, infographics, Web development and programming.” (27)

Research conducted in Germany also suggests that data-driven journalism is mostly teamwork and journalists work alone only occasionally on small projects. Statements by journalists like “Nobody can do everything, so we have to work in a team („Alles kann ja eh keiner, also müssen wir im Team arbeiten.“) or “Data journalism is much more teamwork than traditional journalism”<sup>5</sup> support this claim (13). In contrast, a study at the NYT newsroom shows that “interviewees characterized the collaboration not so much as teamwork but rather as a “back and forth” process between journalist/editor and designer, who discuss the different draft stages of the graphic, from macro to micro concept, before at last the programmer starts his or her work and finalizes the visualization” (30).

Other research suggests a similar approach: there are two ways of data driven journalism. On the one hand, there is data journalism that is produced by one person only during a “normal” day, and on the other hand there are investigative projects, where collaboration of interdisciplinary teams is key (14). Only news outlets with the available resources hire special data journalists (14). “The size of news organizations and management support also influence the production processes. The large broadcasting organization we researched has a separate graphics department. In smaller newsrooms there is a larger demand for multi-skilled journalists who can produce their own information visualizations” (38). “Data journalism teams in news organisations tend to be small, usually only two to five people, even in big companies. Just three people work in the Guardian’s data-journalism team, for example, and this is a news organisation that boasts a daily print readership of about 260,000, and oversees an online version that’s said to be the second-largest English-language online newspaper after The New York Times. The teams’ boundaries are flexible though, and oftentimes reporters from other departments participate in data-processing and analysis. Also, the graphics department may play a role by creating the graphics for the story, or helping out with visualisations. Typically, the data-journalism team works closely with the newsroom and reporters. Indeed, in many cases, data-journalism teams are physically located in the newsroom, rather than in IT departments” (27). “Being in the newsroom really matters” (39). This provides “critical access to editors and reporters as

---

<sup>5</sup> Orig.: „Datenjournalismus ist wesentlich mehr Teamarbeit als traditioneller Journalismus.“

they and the data team develop and plan stories based on data” (39). “Putting developers and reporters in the same area of the newsroom also facilitates another component of these data-journalism teams: ideas for stories that come from either the developers/data miners or the journalists” (39).

Mutual support is essential in data-driven journalism. It is reported that technical problems or questions concerning content are shared on platforms such as Twitter and at least three experts are willing to help (14). Twitter is reported to be a central dissemination and feedback channel where the hashtag #ddj is used to connect people interested in the subject (14).

Besides collaboration between journalists and programmers, also collaboration with the audience is common in data-driven journalism (24). The audience can, for example, offer valuable support in gathering data, which is then called crowdsourcing (24, 31). “The Guardian’s readers are encouraged to help analyze data sets, provide feedback or additional data, formulate research questions, and submit applications and visualizations that they have created from accessible data in the Data Store” (18). As previously mentioned, also in Sweden, crowdsourcing is used quite often to “collect data from the audience or engage them in a topic” (08). “Significant audience engagement with journalism is a relatively new phenomenon” (15). In some cases, even members of competing newsrooms work together on a project. “For instance, New York Times and Washington Post developers are teaming up to create an open election database” (24).

When experts from different fields collaborate communication can sometimes be difficult as “different backgrounds and cultures, insufficient knowledge of each other’s expertise and the lack of any common language often hinder fruitful collaboration” (38). Further, “the fact that journalists and designers do not speak the same language hinders a successful working collaboration. A shift in mindset with journalists thinking more visually and designers taking a more journalistic attitude stimulates collaboration. Nevertheless, this need for a more collaborative mindset is currently hindered by perceived hierarchical differences at the newsroom” (38). Another crucial aspect is the location of data-journalism teams.

#### 3.4.2.7 Publishing the Product

The time of publication is especially important for elaborate projects. A prominent display on the welcome page of the medium as well as sharing it via social media is important (14). In

addition, the placement on the website should not be underestimated (11). A prominent display on the start page or the resort start page not only raises the page views but also is an appreciation of the work of the data team (14). A prominent setup is important as pageviews, retention time and number of viewers are central metrics for measuring the success of data driven projects (14).

As previously mentioned, careful attention has to be paid on the form of the visualization. “Journalists focus less on giving the readers access to “raw” data. They choose design elements that support linear narrative rather than free exploration. Examples of this are the use of timelines, maps, writing, sound and video. These forms are simple, relatively quick to make and limited in functionality. Graphics, lists, tables, grids, searching and filtering are elements that demand more skills and time. One explanation for this trend can be that the advanced journalistic artefacts, as for instance the drillable dataset, do not get enough positive feedback from the readers to make it worth the (considerably lengthy) time it takes to create it. A positive take on this is that the audience needs time to understand and appreciate these new forms of online journalism; that new genres will develop over time, which have a useful balance between data and story.” (03)

Online journalism in general facilitates interactivity and engagement. In data journalism, after publishing the projects, the data teams often also publish the raw datasets, their programmes as well as reports (14). Many journalists publish their data after finishing a project “so that the readers can do their own research on how an issue affects their personal situation.” (06) “Data-driven journalists could share their code and data on open source repositories like GitHub for others to inspect, replicate, and extend. [This is already happening at ProPublica and other outlets.]” (24) The sharing of data is suggested to increase the engagement of the audience and adds credibility and transparency (15). Especially interested and engaged readers have the chance to check the data sets for themselves and find out if they come to the same conclusions as the authors of the story (15, 27). “Therefore, the simple act of publishing the data sets can benefit society in a sense broader than just publishing a news story. Others can find relevant information that goes unnoticed by either journalists or the news stories they produce.” (27)

## **Evaluation**

In the evaluation of the project, quality can not only be measured by its content and its success by the audience but also by its effort, level of complexity of the data and technology

as well as the enhancement of visualization (14). Sometimes, due to lack of resources, the postprocessing of projects is sometimes not possible in a sophisticated way (14).

Knowledge about the audience is important. It is important to know in what way the audience interacts with the story. Most journalists are aware of what works: “People click on cats, after all” (11). Through tracking audiences, journalist can find out more. Some prefer tracking shared content over pageviews (11).

## 3.7 Backgrounds of Data Journalists

### 3.7.1 Educational Background

Data journalists come from various backgrounds. There is “no readily generalizable ‘data journalism’ career path” (11). “Many data journalists began as politics or business reporters and gradually picked up data skills as they became useful to particular stories” (11). Social Sciences and IT-related degrees are quite common among data journalists. Many have come into contact with organizations like the National Institute for Computer-Assisted Reporting (NICAR) and the Organization for Investigative Reporters and Editors (IRE) (11). Those with a background in journalism commonly take courses and trainings to learn technical skills necessary for data journalism (03). Others have no background in journalism but come from a vast range of disciplines. Backgrounds in programming and software development are common. They, in contrast to their peers with a background in journalism, sometimes undertake courses in journalism or obtain a degree (06). However, many report their skills as being “self-taught” as well (33). “Most of them do not share the statistical culture of computer-assisted reporters and have no connection with social science. Relying on some normative principles shared by open source communities and open government advocacy, they have brought in epistemological propositions of how data can support investigative journalism” (06). The literature also reports other degrees or backgrounds of data journalists like East Asian Studies or Urban Studies (12), Art & Design, Anthropology, English, History, Urban Planning, Rhetoric (33).

But how do those with a technical background come into contact with journalism in the first place? One way is via open government meetings or conferences (06). In addition, supportive foundations such as the Sunlight Foundation or the Knight Foundation “have largely funded news projects carried by programmers (Lewis, 2011)” (06). Also universities started offering “journalistic trainings for programmers, such as Medill School of Journalism”

so that “a handful of individuals from the computer worlds have entered the journalistic world in Chicago” (06). Also other universities joined this development as “[w]hen it comes to skills, the training necessary to undertake computational journalism is different from what has traditionally been on offer at journalism schools” (03).

### 3.7.2 „Computational Journalism has no textbooks yet“ - Educational initiatives

„Basic courses in digital journalism, with a focus on social media, are already part of the curriculum at J-schools. The question remains, however, what ‚basic‘ means; having ‚knowledge‘ of software, platforms and techniques does not imply that students need operational skills as well” (20). The importance of integrating these skills in all courses is stressed. Further, “more advanced courses (editing, data journalism, programming, design) for students who want to become experts in these specific areas“ should be offered by respective institutions” (20). As a consequence, more and more universities are offering specialized programs or classes in computational journalism.

Topics discussed in such programs and classes include (31):

- web programming
- SQL
- text/data mining (NLP)
- social computing
- development/deployment for the cloud
- journalistic practices in the digital age
- visualization
- structured journalism & knowledge representation
- network analysis
- computer security
- surveillance & censorship
- web design
- database design
- data journalism and investigative reporting

The courses offered at the different universities and schools vary but “the general idea of computational journalism being an intersection of computing and journalism that requires both technical and journalistic skills is rooted in all of them” (31). In 2011 Jonathan Stray,

Interactive Technology Editor at the Associated Press, wrote in a blog post: “‘Computational journalism’ has no textbooks yet”, but provides a good outline of what one could include in the list (Stray 2011a)“ (31).

The following list provides a short overview of already existing educational opportunities (the focus of the literature lies on the US):

University	Program/course	What/how	Since	Paper
The Georgia Institute of Technology	Course in Computation and Journalism	research in the field, “ the academic approach to CEJ is now widely known under the term ‘computational journalism’. The field surfaced at the Georgia Institute of Technology in 2006, and gained international acceptance as it was adopted by a number of science communities in the USA and in Europe, and also by funders such as the Knight Foundation“	2006	18, 34
Missouri School of Journalism	NICAR	“how to treat data as a source”; most important training institution for journalists to use information technology; teaches students how to apply data science to all the news that’s fit to print	1989	24
Northwestern University’s Medill School	Knight Lab	Classes on enterprise reporting with data, interactive storytelling with JavaScript; Scholarships for programmers; “pairing journalism and computer science students together to develop interactive projects“		24, 39, 34, 12
University of California Berkeley	free online data journalism training	“10 courses funded by the Kickstarter campaign were taught by leading practitioners in the field. “For Journalism” endures as a free online resource for anyone who wants to learn more, including webinars, ebooks, code repositories, and forums“	2013	24
Cardiff University	Master program	will introduce master in computational journalism		24
Columbia University Graduate School of Journalism	Lede Program	post-baccalaureate certification course that offers training in data, code, and algorithms to journalists; “will equip students with the technical skills required to enroll in the dual Journalism/Computer Science master’s program that Columbia began to offer in 2010“	2010	24, 34
Columbia University Graduate School of Journalism	Master program	dual Journalism/Computer Science master’s program	2010	24
Temple University	Course	“new course by computer scientist-turned-reporter Meredith Broussard”; essential reporting skills,		24

		critical thinking, multimedia storytelling skills, visual analysis, Journalism**, Vox, <a href="http://FiveThirtyEight.com">FiveThirtyEight.com</a>		
University of Miami	visualization classes	“Visualization classes are now part of the core program for undergraduate journalism majors and in the university’s master’s degree program, along with mandatory introductions to design and Web design”		24
Stanford University	Computational Journalism Lab	Classes on relational data, basic statistics, investigative reporting tools, mapping		24

But also engaged media organizations train their staff. The online editors and reporters of BBC News, for example, are „required to attend conferences and take in-house trainings to learn to use basic spreadsheet applications such as Excel and Google Docs“ (39).

### 3.7.3 Necessary Skills for Data Journalism

Back in 1991 Meyer already claimed that “a journalist has to be a database manager, a data processor, and a data analyst” (Meyer, 1991:1). “Data journalism adds new requirements to a journalist’s profession. Now they need skills for data analysis, web development and community management, as readers are invited to participate in the stories. However, the most important factor for successful data journalism is the journalistic sense of what is relevant and interesting, and what questions need to be asked. Only after these are determined can data’s usefulness kick in to address those questions” (27). Along those lines, media organizations need to figure out, who is equipped with necessary skills to handle those tasks. “The question of the skills required to produce data journalism seems particularly crucial, with findings underlining the need to master skills at the intersection between journalism and technology (Royal, 2010; Karlsen and Stavelin, 2014)” (28).

Certain skills are considered to be crucial for data journalism. “When looking toward the needs and requirements of the professional field compared to the ‘traditional’ journalist, we found new and diverse professional profiles: Artistically skilled journalists who can serve the multi-disciplinary requirements of interactive information graphics almost autonomously, even though those persons are hard to find” (12). However, “[t]he technical skills are subordinated to the unbroken tradition of journalism” (03). Journalistic values are still a key element in data journalism (03). “The technical skills are subordinated to the unbroken tradition of journalism” (03). Along these lines, computational journalists and data journalists “need(s) to master both the inverted pyramid structure of journalistic storytelling and basic iteration statements found in any programming language such as the ‘while’, ‘for each’ and

‘for’ statements” (03). It is argued that data journalists should at least have minimal data-related skills (11), or in other words “newspapers should set up dedicated units staffed with people familiar with journalism as much as code” (06). In the literature three main competences for data journalists are mentioned (14):

- 1) data analysis/programming skills
- 2) design
- 3) wording skills

The intensity of needed data education seems to vary however: “In IDJ<sup>6</sup>, data skills are used at advanced levels, including coding. This could be done in team work. In GDJ<sup>7</sup>, the data skills are at the basic level, for example, the capability of using Excel and some data visualization and analyzing tools” (19, similar idea in 28).

Important characteristics and skills of a data journalist according to the literature include (amongst others):

- Curiosity (03, 33)
- Logic of investigative reporting (03)
- Sense of logic (03)
- Technical attitude (33)
- Eager to learn new skills (33)
- Problem-solving skills (03, 33)
- Best: coding/programming; languages like Python, Ruby, PHP (06, 33, 38, 39)
- Good understanding of the Web (all 33)
- “skilled in obtaining all types of computerized information” (36)
- Data mining (38, 39)
- Writing (38)
- Designing (38)
- Data-related skills (gathering, analyzing, etc.) (06, 11, 39)
- Understanding of numbers (13)
- Understanding of statistical and social science methods (13, 14)
- Computational thinking (18)
- Identifying news (39)

---

<sup>6</sup> Investigative Data Journalism (IDJ)

<sup>7</sup> General Data Journalism (GDJ)

- “proficient in multimedia, interactive news, computer-assisted reporting and graphics” (39)
- Software project management (14)
- Cartography (14)

However, in the literature it is also suggested that not a single person must inherit all of those traits (38), often it can be important to know “one person, who can code, and that he or she can ask for help” (19). “Our empirical research shows that forerunners in the use of information visualization often work with multi skilled professionals. However, as soon as the production of information visualizations becomes routinized, the need is felt for a more diverse team of professionals, such as programmers, illustrators and designers. This is especially the case with larger newsrooms. Our research also shows the importance of transformer skills. Journalists and designers often lack this capability of converting journalistic content into a visual form. This skill, however, seems decisive for smooth collaboration.” (38)

Generally it seems to be important to hire “reporters and developers who bridge the skills gap: Data-journalism team members who have backgrounds and skills in journalism and data mining, development and/or coding are valuable assets to data-journalism teams because they can sort and analyze data and then identify news and trends that resonate with their audience” (39). “For their part, developers, coders and data miners should have an understanding of best journalistic practices: accuracy, sourcing, fact-checking, creating narratives and writing stories” (39). When working with datasets, it is necessary for journalists to understand the “‘reliability’ and ‘validity’ of the data that they use in their own stories” (23).

“Aron Pilhofer, editor of interactive news at The New York Times, says that as long we continue to talk about the skills gap between journalists and developers, the gap will persist. He says investments must be made to make journalists comfortable with basic data skills (spreadsheets, data analysis), even if it means hiring a strong reporter who is willing to use new tools over a journalist who is a stronger writer.” (39)

“Kirk suggests that successful information visualization requires eight capabilities. The ‘eight hats of data visualization design’ are (1) the initiator, (2) the data scientist, (3) the journalist, (4) the computer scientist, (5) the designer, (6) the cognitive scientist, (7) the communicator

and (8) the project manager (Kirk 2012). However, he does not take a clear stance on whether one or several persons should master these capabilities” (38).

## 3.8 Enabling and Hindering Factors of Data-driven Journalism

### 3.8.1 Accessibility

Even where access to data is, in principle, possible a central factor is accessibility. This accounts both for the journalist, who has to go through the material to find the story, as well as for the reader, who wants to take a closer look at it. A prototypical example for this is the publication of data in a non-machine readable format like PDF. “Furthermore, data sets almost always include errors, which journalists have to catch (1). Particularly when scraping data, the errors can occur in data sets” (27). “Much of the time, government data is often “dirty,” with missing metadata, incorrect fields, or gaps in collection. Journalists have to extract data from PDFs, validate it, and clean up data sets to make them usable in applications, report it out, and then present it in context” (24). Most Government data in Britain is published in PDF format which poses problems for analyzing the data further. Nevertheless The Guardian and the New York Times have tackled this problem and provided tools in order to open up the possibility to explore this data in an interactive way (01).

### 3.8.2 Collaboration and Position in Newsroom

“Collaboration as a crucial condition to produce databased visualizations is closely related to the key factor attitude. Important steps in the production process of a data-based visual are: data retrieval, analyzing and filtering data sets, extracting the message from the data, transforming the message into a story, visualizing the story, and programming the story” (12). Data journalism is considered to be much more of a team effort than traditional journalism (13). Numerous experts help to produce the final piece (14). As mentioned previously (see Collaboration), (cross-disciplinary) collaboration and mutual support are key in data journalism and happen also between different media outlets (14, 18). However, “[o]ften there is a physical divide between data-journalism teams and the news desk” (39), which is often considered as a hindering factor.

### 3.8.3 Tools

Data journalism generally profits from the possibilities of new media. Free online tools are seen as helpful, for example ManyEyes (1). “The ‘Document Cloud’ tool, which uses cloud technology, was developed in cooperation between the group at the New York Times and the non-profit news organization, ProPublica. Its purpose is to help journalists process and publish documents containing unstructured data. The collaborative approach to developing the tool is related to ideas that are central to computer science” (08). Numerous free and user-friendly tools are available for journalists working with data (19). (See also: Building and visualizing).

### 3.8.4 Supportive Infrastructure

The National Institute for Computer-Assisted Reporting (NICAR) has proven to be a main supportive infrastructure for the development of data journalism. Founded in 1989 by the Organization for Investigative Reporters and Editors (IRE) and supported by the Missouri School of Journalism it offers data training for journalists (06, 09, 22, 30). The development of data journalism is also fueled by communities outside educational spectrum like, for example, Hacks & Hackers (14, 22, 35), Sunlight Foundation or the Knight Foundation (09).

In addition, the genre is acknowledged by the rise of awards for data journalism like The Data Journalism Award, as well as the founding of new categories within already existing awards. The Amnesty International Media Awards now include the category “Digital Innovation” while the British Press Awards the “Digital Award” is awarded. This leads to a growing recognition of the field (21).

### 3.8.5 Lack of Resources: Time and Money

Particularly at daily newspapers “time is precious for journalists” (1) and stories have to be produced fast. However, data stories often need more time, especially if bigger data sets or deeper investigations are involved, which have to be cleaned, analyzed and visualized. “[I]n news coverage time is the limiting factor for data visualization, because visuals, e.g. interactive graphics, cannot be produced in a few hours like an article; data visualization requires time. Therefore, reporting new information first or getting the scoop cannot be the aim of data journalism; accuracy always comes before speed” (12). Data work is more time consuming: “Extracting meaningful information from forms is among the most expensive and

time-consuming large news investigations. Its cost sometimes results in abandoning promising stories” (34).

Another big constraining factor is money. „Many media companies still hesitate to invest in developing specialized data units within their staff, because hiring ready-made specialists or training existing ones seems too expensive” (26). In addition, the return on investment „seems feeble, uncertain, or painfully fruitless“ (28), as data journalism often requires a lot of time with no immediate results (28). However, the evolution of journalism and the gaining importance of (big) data makes it important for publishers to consider the introduction of data journalism (26). Data stories are reported to offer advantages compared to normal news stories. Some research finds that data stories live longer. Readers tend to come back to a data story even months after the original publishing (15). Also, readers are said to spend more time on data stories (27).

Numerous studies found that the bigger the newspaper the more likely it is to invest in special data journalism departments and hire experts. „Larger organizations were more likely to undertake data work that involved a division of labor, with computer-assisted reporters, graphic designers, statisticians, and programmers working on teams. Small organizations were more likely to have “one-man band” who acquire data skills as needed due to their own initiative“ (11).

### 3.8.6 Business Models & Revenue

As previously mentioned, a hindering factor of data journalism could be the fact that there “is no proven business model for data journalism [yet]. Overall, news organizations conclude that data journalism is so far a matter of journalistic value, and hold potential for impact rather than revenue” as data journalism often consumes more time than “normal” journalism (27). However, research into the topic suggests different possibilities for seemingly profitable business models (27):

- Premium model: making people pay for better visualizations
- Go-to store for data: organizations as well as individual citizens. Access could be a paid service
- Service: analysis of data analysing for organisations and companies

While the news industry in large part is still struggling, nonprofit organizations and other startups already embraced data-driven work. Prominent examples include ProPublica or the Texas Tribune. They are not concerned about revenue: “The Tribune, which has emerged as a bright spot in the firmament of online media for state government, focuses on covering the Texas statehouse. It is now one of the most important examples of data journalism in the United States, given the success of its data visualizations and interactives” (24).

Other examples that “enjoy healthy growth rates and sometimes impressive profits” (27) include, for example, Bloomberg. The company, delivering financial data to its users, has hired numerous journalists and bought the struggling Business Week. The Economist has also proved to be a big player in the data business. “The magazine has built an excellent, influential brand on its media side. At the same time the “Economist Intelligence Unit” is now more like a consultancy, reporting about relevant trends and forecasts for almost any country in the world. They are employing hundreds of journalists and claim to serve about 1.5 million customers worldwide” (27).

“News organisations are looking for sustainable ways to support data journalism, but viable business models have yet to be discovered. News organisations have visions of becoming number-one data stores and of adding the provision of data analysis to their oeuvre of paid service offerings. The most likely future scenario will see news organisations existing in a state of symbiosis with other data service providers, booming in the new ecosystem evolving around this area” (27).

### 3.9 Gender aspects in Data Journalism

Gender is also mentioned in some of the research papers, often related to diversity in data journalism teams: “Each member of the team expressed a sincere sensitivity to the issues and a recognition of the importance of having a diverse team. However, they were unclear how to overcome the dearth of women who possessed the skills or had an interest in programming.” (33)

“More diverse newsrooms will produce better data journalism. Diversity has been a challenge in the media for decades. Although far more minorities and women work in professional journalism than a century ago, a 2013 survey of American Society of News Editors (ASNE) found that of the 38,000 journalists currently working at 1,400 U.S. newspapers, 4,700 are minorities.” (24).

Focusing on women, challenges are led back to “broader issues with women in technology” (24) or “an issue in the tech world” (33) as well as “societal issues” (24) on a structural level which result in a lack of women.

“Several indicated that they often work with female multimedia producers or reporters in developing the interactives, so the female perspective is not completely absent.” (33).

“Research suggests that teams with both men and women on them are more profitable and innovative. According to the National Center for Women and Information Technology, mixed gender teams produced information technology patents that are cited 26 percent to 35 percent more often than the norm” (24).

Furthermore gender-related issues in the papers are related to interest and experience: “The opportunities exist in being able to present technology tools and concepts to a female audience, thus increasing the representation of women in this field. The challenge will be in gaining females’ interest in such an activity, with the hope being that as it relates to a communications context, women may be more likely to gravitate toward it or to want to be associated” (33).

One paper covered among other things the “perceived experience of working with data journalism methods and gender [...] The results indicate that the males responders tend to describe themselves as more experienced than the woman” (08).

### 3.10 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 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 actually integrated into different organizations. Along those lines, (Appelgren and Nygren, 2014), who themselves focused on the analysis of data

journalism in Sweden, 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 of time. 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 resources discrepancies between small and large newspapers led them to tell different stories (scope, depth, 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), on the other hand, 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, the majority of authors and author collectives within our corpus did not discuss research gaps or suggest paths for further research.

## 4 Conclusions & Outlook

In this document we report on the analysis of selected research literature on data journalism and related areas. We looked at both the structure and selected topics of published research in the field. Following the rigid method of a structured literature review, we selected a corpus of publications, which were subsequently examined using computational methods and qualitative software-assisted content analysis.

Both data journalism and its accompanying (more or less critical) research have been developing rapidly in the past two decades. The last three years in particular 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.

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 concepts. Many publications just report what has been investigated without giving (much) space to theoretical or methodological questions. 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). We would thus 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; Peters, 2001; Schudson, 2001, 2010). For instance, capturing concepts of data journalism through a history-of-ideas lens could be very enriching.

Another issue is the lack in variety of research methods. While many research projects employ two different qualitative methods or triangulation (Flick, 2009), 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), 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, we also see that questions of gender are nothing much addressed in data journalism research. Women seem to be a minority in data journalism, however, gender issues have to be analysed on several levels. 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 and meaningful 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 – the paper 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.

Regarding user needs and requirements we also hint to user research which has been conducted in the course of the VALiD project. In accordance with some works of our research literature it revealed that “because of limited statistical and programming skills, the developed tool have to be very easy-to-use and self-explaining.” (Niederer e.a. 2016)

## 5 References

Aitamurto T, Sirkkunen E and Lehtonen P (2011) *Trends in data journalism*. Available from: [http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B\\_Hyperlocal\\_Trends\\_In%20Data\\_Journalism.pdf](http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf) (accessed 31 August 2015).

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 (2012) 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*, Studies in the Digital Humanities, Sheffield: HRI Online Publications.

Appelgren E and Nygren G (2014) 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, et al. (eds), *Digitale Methoden in der Kommunikationswissenschaft*, Berlin: Digital Communication Research, pp. 87–111.

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.

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

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377–387. <https://doi.org/10.1145/362384.362685>

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, Li C, Yang J, et al. (2011) Computational journalism: A call to arms to database researchers. In: *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research (CIDR 2011)*, pp. 148–151.

Cohen S, Hamilton JT and Turner F (2011) 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.

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

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

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

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 Association for Education in Journalism and Mass Communication (AEJMC) National Convention*, Phoenix, AZ.

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. *Nick Diakopoulos - Musings on Media*, Blog. Available from: <http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/> (accessed 31 August 2015).
- Diakopoulos N (2012) *Cultivating the landscape of innovation in computational journalism*. New York, NY: CUNY Graduate School of Journalism & Tow-Knight Center for Entrepreneurial Journalism. Available from: [http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos\\_whitepaper\\_systemicinnovation.pdf](http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos_whitepaper_systemicinnovation.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 from: [http://mediapusher.eu/datadrivenjournalism/pdf/ddj\\_paper\\_final.pdf](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): 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: McCallum K (ed.), *Media, Democracy and Change: Refereed Proceedings of the Australian and New Zealand Communication Association Annual Conference*, Canberra.
- Flew T, Spurgeon C, Daniel A, et al. (2012) The promise of computational journalism. *Journalism Practice* 6(2): 157–171.
- Flick U (2009) *An introduction to qualitative research*. 4th ed. London: Sage.
- Garrison B (1999) Newspaper size as a factor in use of computer-assisted reporting. *Newspaper Research Journal* 20(3).
- 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: 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: *Center For Advanced Study in the Behavioral Sciences Summer Workshop*, Palo Alto. Available from: <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 from: <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, CSCW '15*, New York, NY: ACM, pp. 1170–1175.

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).

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 from: <https://github.com/ckreibich/scholar.py> (accessed 15 April 2016).

Lecheler S and Kruike-meier 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 and 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.

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.

Lorenz M (2010) Status and outlook for data-driven journalism. In: European Journalism Centre (ed.), *Data-driven journalism: What is there to learn?*, Amsterdam, pp. 8–17. Available from: [http://mediapusher.eu/datadrivenjournalism/pdf/ddj\\_paper\\_final.pdf](http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf).

Lugo-Ocando J and Brandão RF (2015) Stabbing news: Articulating crime statistics in the newsroom. *Journalism Practice* 9: 1–15.

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).

Meyer P (2002) *Precision journalism: A reporter's introduction to social science methods*. 4th ed. 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.

Peters JD (2001) 'The only proper scale of representation': The politics of statistics and stories. *Political Communication* 18(4): 433–449.

QSR International (2015) *Nvivo*. Available from: <http://www.qsrinternational.com/>.

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: MIT Press.

- Rogers S (2008) Turning official figures into understandable graphics, at the press of a button. *Inside the Guardian Blog*. Available from: <http://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata> (accessed 21 June 2016).
- Royal C (2012) The journalist as programmer: A case study of the New York Times Interactive News Technology Department. *#ISOJ The Official Research Journal of the International Symposium on Online Journalism* 2(1).
- Scholl A (2011) Der unauflösbare Zusammenhang von Fragestellung, Theorie und Methode. Die reflexive Bedeutung der Methodologie (nicht nur) in der Journalismusforschung. In: Jandura O, Quandt T, and Vogelgesang J (eds), *Methoden der Journalismusforschung*, Wiesbaden: Verlag für Sozialwissenschaften, pp. 15–32.
- Schreier M (2012) *Qualitative content analysis in practice*. Thousand Oaks: Sage.
- Schudson M (2001) The objectivity norm in American journalism. *Journalism* 2(2): 149–170.
- Schudson M (2005) Four approaches to the sociology of news. 4th ed. In: Curran J and Gurevitch M (eds), *Mass Media and Society*, 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, Haan Y de and Buijs L (2014) Visualizing news: Make it work. *Digital Journalism* 2(3): 344–354.
- Stavelin E (2013) Computational journalism: When journalism meets programming. Dissertation, Bergen: University of Bergen.
- 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 Jr. EC and Oh S-K (2015) Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism Studies*. 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.
- Uskali T and Appelgren E (2015) The Third Nordic Data Journalism Conference: Call for papers. Available from: <http://www.noda2016.fi/callforpapers.html> (accessed 8 July 2016).
- 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)*, Montpellier: IEEE, pp. 349–356.

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 from: [http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english\\_0.pdf](http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english_0.pdf).

## 6 Appendices

### 6.1 List of definitions for data journalism

Author	Defined Term	Definition
Felle (2015)	Digital data reporting, CAR	<p>Digital data reporting, loosely defined, acts both as an investigative and a storytelling tool for journalists.</p> <p>While CAR journalists pioneered the use of dataset inquiry, some respondents argued, digital data reporters are conducting significantly more complex investigations into far larger datasets and presenting the results of such investigations in far more compelling ways using visualisation and interactivity tools. The method of investigation has not fundamentally changed but the complexity of the work has.</p> <p>It should also be noted that CAR was always an investigative tool, never a storytelling tool. In digital data reporting, however, these two cannot be considered separately – both the investigative method and its ensuing visual and interactive storytelling allow for greater accountability.</p>
Coddington (2015)	Computational journalism	<p>Computational journalism has at times been used by scholars to include CAR and data journalism, conflating the previous two forms; indeed, the most common definition of computational journalism seems to encompass both CAR and data journalism: “the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism” (Hamilton and Turner 2009, 2). But by defining them so broadly, this definition does not allow much room to draw significant distinctions between each of the three practices. Instead, following Diakopoulos (2011), I define computational journalism here as a strand of technologically oriented journalism centered on the application of computing and computational thinking to the practices of information gathering, sense-making, and information presentation,</p>

rather than the journalistic use of data or social science methods more generally.

Cohen et al. (2011) computational journalism

“A new field is emerging to promote the process: computational journalism. Broadly defined, it can involve changing how stories are discovered, presented, aggregated, monetized, and archived. Computation can advance journalism by drawing on innovations in topic detection, video analysis, personalization, aggregation, visualization, and sensemaking.<sup>6–8,13</sup>”

Daniel & Flew (2010) Computational journalism

“Computational journalism can be broadly defined as the application of computer science techniques to the activities of journalism. Its foundation lies in computer assisted reporting techniques and its importance is increasing due to the: (a) increasing availability of large scale government datasets for scrutiny; (b) declining cost, increasing power and ease of use of data mining and filtering software; and Web 2.0; and (c) explosion of online public engagement and opinion.”

“Computational journalism can be broadly defined as the application of computer science techniques to the activities of journalism. It is defined more formally by Hamilton and Turner (2009: 2) as: the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism. In some ways computational journalism builds on two familiar approaches, computer-assisted reporting (CAR) and the use of social science tools in journalism. Like these models, computational journalism aims to enable reporters to explore increasingly large amounts of structured and unstructured information as they search for stories.”

“Computational journalism involves using tools to access, manipulate and present data and report news. By the use of emerging software across vast datasets, journalists may identify clues that lead to original investigations and enhanced reportage.”

Diakopoulos (2011) computational journalism

“the application of computing and computational thinking to the activities of journalism including information gathering, organization and sensemaking, communication and presentation, and dissemination and public response to news information.”

Diakopoulos (2015) computational journalism

In this paper, I proffer a new branch of research in computational journalism that inverts the typical

		<p>tool-orientation and foregrounds journalism by making computation its object.</p>
Hannaford (2015)	Computational journalism	<p>“Others have described the way in which CAR has evolved into Computational Journalism by utilising Web 2.0 technology which allows for new forms of journalistic storytelling such as multimedia, interactive features as well as journalistic investigation (Flew et al., 2012); Hamilton &amp; Turner, 2009; McAdams, 2014). The significant implication here is that whereas CAR (and data journalism) was largely carried out by people who solely identified as journalists, computational journalism brings technologists and journalists together to develop new tools and interactive storytelling techniques (Flew et al., 2012; Lewis &amp; Usher, 2014). Sometimes the end product will not be a traditional story, but a tool that allows the audience to use the data to answer questions of personal interest (Parasie &amp; Dagiral, 2012).</p>
Karlsen & Stavelin (2014)	Computational journalism	<p>“When asked to define computational journalism the respondents did not offer narrow or distinct definitions. They all gave wide and open definitions that included many different activities and forms. One of the respondents said that he “uses computational journalism for everything from research and fetching data to visualizations” and that computational journalism is to “find new ways to both find stories, and to tell stories”. Another respondent offered a similar wide definition, “computational journalism is everything from the simplest use of Excel to heavy tools that enable journalism that is impossible without these tools”. They emphasized that computational journalism empowers them to do more. Analysis of data is central to this. They were able to “analyse large datasets and find support for existing theories, or find new truths, new trends”.”</p>
Turner and Hamilton, 2009	Computational journalism	<p>For now though, we define computational journalism as the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism. In some ways computational journalism builds on two familiar approaches, computer-assisted reporting (CAR) and the use of social science tools in journalism championed by Phil Meyer in Precision Journalism: A Reporter’s Introduction to Social Science Methods (Rowman and Littlefield, 2002). Like these models, computational journalism aims to enable reporters to explore increasingly large amounts</p>

		of structured and unstructured information as they search for stories.
Young & Hermida (2015)	Computational journalism	we define computational journalism using Diakopoulos (2011, 1, emphasis in original) as “the application of computing and computational thinking to the activities of journalism including information gathering, organization and sensemaking, communication and presentation, and dissemination and public response to news information.”
Gynnild (2014)	Computational journalism, CAR	<p>“The umbrella term computational exploration in journalism (CEJ), embraces the multifaceted development of algorithms, data, and social science methods in reporting and storytelling. CEJ typically involves the journalistic co-creation of quantitative news projects that transcend geographical, disciplinary, and linguistic boundaries.”</p> <p>“I define computational exploration in journalism as ‘the innovative processing that occurs at the intersection between journalism and data technology’. The concept embraces the experimental use of algorithms, data, and social science methods in the news media, a process that ranges from data retrieval and data analysis to data visualization. CEJ may include the building of new technological tools for data mining and visualizations, or it may imply using existing tools in new ways. Computational exploration in journalism transcends former geographical, disciplinary, and linguistic boundaries and is carried out at individual and group levels as well as at organizational levels.”</p> <p>“CAR includes techniques such as data searches on the web, spreadsheet and/or statistical analysis of various public records, and geographical and other information mapping. CAR also allows users to seek background information online and to interview people by email or social media. The term ‘computer-assisted reporting’ emerged long before digitalization of the media.”</p> <p>“CAR is mainly associated with investigative reporting, a small branch of journalism. Some people in the industry tend to speak of CAR as the earlier generation of data devoted journalists.”</p> <p>“Hamilton and Turner (2009: 1) define computational journalism as ‘the combination of algorithms, data and knowledge from the social sciences to supplement the accountability function of journalism’. By contrast, Nick</p>

Diakopoulos has chosen a work process approach and defines computational journalism as: ... the application of computing and computational thinking to the activities of journalism including information gathering, organization and sensemaking, communication and presentation, and dissemination and public response to news information, all while upholding core values of journalism such as accuracy and verifiability. It is inclusive of CAR (ComputerAssisted Reporting) but distinctive in its focus on the processing capabilities (e.g. aggregating, relating, correlating, abstracting) of the computer in comparison to mundane aspects of storage or access. The field draws on technical sub-fields of computer science including information retrieval, artificial intelligence, content analysis, visualization, personalization, and recommender systems as well as aspects of social computing and information science. (2011[2010]: 1)”

Davenport et al. (1996) computer assisted journalism, computer assisted reporting

“But, what exactly does computer-assisted mean in journalism? For some people, the term is a general one that simply means using a computer to produce a news product. To others, like Hill, it is the skill of using a computer in all ways -from CO-ROMs to electronic morgues- to gather information to write a news story. And, still to others, the phrase means a very specific skill, such as the method of getting information off a commercial online database service such as Lexis/Nexis or the procedure of analyzing data using spreadsheet software.”

“Kathleen Hansen introduced the term computer-assisted reporting in the results of their study to mean the additional use of personal computers or creating databases and data generally. They didn’t dwell on a definition, but Margaret DeFleur and Lucinda Davenport did.<sup>6</sup> I The phrase computer-assisted reporting is a reference to all computer gathering skills, not JUST to commercial online databases. Their definition included a variety of computer-information gathering skills that journalists need in order to: 1) search for information in commercial online databases from publicly available online databases within electronic bulletin board services (BBS’s) or from a newspaper’s own electronic morgue; 2) analyze public agencies’ electronic records; and 3) build their own customized topical databases. Thus, the phrase computer-assisted reporting is a reference to all computer information-gathering skills, not just to commercial online databases.<sup>7</sup>”

Davenport et al. (2000)	computer-assisted reporting	<p>“The computer-assisted reporting revolution may now be more of an evolution as computers are more routinely used for gathering and processing information. Indeed, the term, “computer-assisted reporting,” may no longer be relevant as reporters use computers for gathering information on a regular basis as part of their writing and reporting process.”</p> <p>“That revolution in electronic news gathering is now being carried almost entirely by the reporting staff. As this study shows, reporters in increasingly more newsrooms have learned to use the Internet, online databases, CD-ROMs and BBS’s. Furthermore, reporters are routinely accessing and analyzing data from such sources as electronic public records and newspaper-developed databases.”</p>
Parasie & Dagiral (2013)	computer-assisted reporting” (CAR)	<p>“computer-assisted reporting” (CAR) has sometimes been used as a label for stories revealing injustice in society by pointing out the existence and the causes of a social issue, and identifying solutions to it.</p> <p>...</p> <p>Most stories reveal public issues on the basis of data analysis, drawing the attention of public opinion and sometimes calling for public intervention.</p>
Aitamurto et al. (2011)	Data journalism	<p>“According to the interviews conducted for this study, data journalism is perceived to be a tool for producing better journalism with the help of statistical methods, visualisation and interactive means of presenting information. Journalists see data journalism as a way to find hidden stories—stories that wouldn’t otherwise get told.”</p>
Appelgren & Nygren (2012)	data journalism	<p>In this paper we choose to use the term “data journalism” and define it as a form of storytelling based on journalistic research in the form of collection, compilation, processing, analysis and presentation of large data sets. This form of journalism is often presented as interactive graphics or seemingly simple interfaces for the user on webpages or in mobile publishing channels. It is usually published in combination with news articles and audio or video content or print, depending on the publishing house.</p>
Appelgren & Nygren 2014	Data journalism	<p>Data journalism is currently an emerging form of storytelling, where traditional journalistic working methods are mixed with data analysis, programming and visualization techniques (Nygren, Appelgren, and Hüttenrauch 2012).</p>
Bounegru	Data journalism	<p>“By enabling anyone to drill down into data sources</p>

(2012, 22)		<p>and find information that is relevant to them, as well as to verify assertions and challenge commonly received assumptions, data journalism effectively represents the mass democratisation of resources, tools, techniques and methodologies that were previously used by specialists — whether investigative reporters, social scientists, statisticians, analysts or other experts.”</p>
<p>Coddington (2015)</p>	<p>data journalism</p>	<p>Sometimes referred to as data-driven journalism, data journalism seems to have taken up the mantle of CAR in contemporary professional journalism. Though it is less preferred by scholars, data journalism appears to be the term of choice in the news industry for journalism based on data analysis and the presentation of such analysis (though note the ambivalence toward the term found by Appelgren and Nygren 2014). Professional definitions have tended to be broad, characterizing data journalism as essentially any activity that deals with data in conjunction with journalistic reporting and editing or toward journalistic ends, as in Stray’s (2011) definition of data journalism as “obtaining, reporting on, curating and publishing data in the public interest.” Several others have defined data journalism in terms of its convergence between several disparate fields and practices, characterizing it as a hybrid form that encompasses statistical analysis, computer science, visualization and web design, and reporting (Bell 2012; Bradshaw 2010; Thibodeaux 2011). Data journalism has also been closely associated with the use and proliferation of open data and open-source tools to analyze and display that data (Gynnild 2014), though open data is not necessarily or exclusively a part of its domain of practice (Parasie and Dagiral 2013).</p>
<p>Fink Anderson (2015)</p>	<p>&amp; Data journalism</p>	<p>“Data journalism is ultimately a deeply contested and simultaneously diffuse term, and thus would seem to impose analytical difficulties for those who wish to study it.”</p>
<p>Knight (2015)</p>	<p>Data journalism</p>	<p>“Data Journalism appears to be the inheritor of two older news practices: infographics and computer-assisted reporting (CAR).”</p> <p>“Data journalism is defined by Simon Rogers (2011) as</p>

'a field combining spreadsheets, graphics data analysis and the biggest news stories' (para.110), while Mirko Lorenz (2010) refers to it as a process that goes from analysing, filtering and visualising data in a form that links to a narrative and is useful to the public. The emphasis on graphics and visualisation is common, and for some observers, data journalism is fundamentally the production of news graphics, and fits within that framework of practice, with elements of design and interactivity taking precedence (Bradshaw 2010; Lorenz 2010; Rogers 2011). For others, the focus on large data sources, often acquired through leaks or freedom of information requests, and the extended and complex analysis of this data is important, linking data journalism to the practice of investigative journalism, as Meyer (1991) did with CAR. For the purpose of this study, I have taken the broadest possible definition of data journalism: a story whose primary source or 'peg' is numeric (rather than anecdotal), or a story which contains a substantial element of data or visualisation. This broad definition allows for the widest possible net, catching as many examples as possible of journalism that incorporate data, in order to create an understanding of the field."

Loosen et al. Data journalism (2015)

These definitional approaches (e.g., Anderson, 2013; Appelgren & Nygren, 2014; Gray et al., 2012; Coddington, 2015; Fink & Anderson, 2015; Gray et al., 2012) highlight the following presumed characteristics of data journalism:

- It builds on (usually large) sets of quantitative (digital) data as 'raw material' which is subjected to some form of (statistical) analysis in order to find stories in it or tell stories with it;
- the results "often need visualization" (Gray et al., 2012: n.p.), i.e. they are presented in the form of maps, bar charts and other graphics;
- it is "characterised by its participatory openness" (Coddington, 2015: 337) and "so-called crowdsourcing" (Appelgren & Nygren, 2014: 394) in that users help with collecting, 2 analysing or interpreting the data – going along with journalists relinquishing their "interpretive authority" (Weinacht & Spiller, 2014: 414; own translation);
- and it follows an open data and open source approach by publishing the raw data a story is built on. However, scholars' definitional approaches are often contradicting: While Anderson (2013: 1005), for instance, places data(-driven) journalism within the broader field of "computational journalism", Coddington (2015) explicitly distinguishes between these two

concepts (as well as computer-assisted reporting). Although he (2015: 333) acknowledges that computational and data journalism “are not mutually exclusive” and that they “inevitably overlap”, he identifies “significant differences between these forms of practice” and makes the point that “not all data journalism is computational” (Coddington 2015: 336). In addition, computational journalism, for Coddington (2015: 337–341), is characterised to a much lesser extent by professional expertise as well as transparency. Consequently, Fink & Anderson (2015: 478) lament the “lack of a shared definition of data journalism”, which (presumably not only) Coddington (2015: 332) considers “fundamental” for building “a coherent body of scholarship”.

Uskali & Kuutti (2015) Data journalism or data-driven journalism “Data journalism or data-driven journalism has often been defined as journalism based on large data sets, otherwise known as ‘big data’ (Rogers 2011; Bounegru et al. 2012). In a similar vein, a data journalist could be defined as a person who creates news stories based on large data sets. In addition, Parasie and Dagiral (2012), when summarizing the evolution of data-driven journalism in Chicago, have used the terms “computer-assisted-reporters” and “programmer-journalists”.

Tabaray et al. (2016) data journalism, computational journalism “Its creation relies on a variety of computer skills needed to collect, process, combine, and visualize data – whether it be numbers, texts, photographs, or audiovisual content available in digital formats. The term most commonly used today to refer to this collection of heterogeneous practices is ‘data journalism’. (Parasie and Dagiral, 2013b: 53, original translation)”

“Data journalism is based on the use of quantitative data, but what interests us here, beyond the simple manipulation of quantitative objects, is the transformation of the journalistic product and its production process. This interest stems from the fact that, according to Desrosières (2008), ‘quantification offers a specific language that provides remarkable properties of transferability, possibilities of standard manipulations through calculation, and routinized interpretation systems’ (p. 12, original translation).”

“Hamilton and Turner (2009) offer the following definition of computational journalism: In some ways computational journalism builds on two familiar approaches, computer-assisted reporting (CAR) and

the use of social science tools in journalism. Like these models, computational journalism aims to enable reporters to explore increasingly large amounts of structured and unstructured information as they search for stories. (p. 2)”

“Flew et al. (2012), however, assert that a more complete definition of computational journalism than given above can be obtained in combination with quantitative methods identified by Philip Meyer. Among others, Meyer (2002) cites statistical analysis, polling, surveys and observation, and collection and interpretation of public data (p. 3). The idea of computer-assisted reporting and research is adopted, and in some ways invented, in English-speaking newsrooms (Thiran, 1996). Enthusiasm for the use of quantitative analysis and codes developed in the social sciences is seen in the concept of ‘precision journalism’ proposed by Meyer in the 1970s (Meyer, 2002). As for the computerization of these forms, the progressive and unequal integration of computers in various print newsrooms since the 1980s must be taken into consideration.”

Howard (2014) Data journalism, data-driven journalism  
 “Although journalists have been using data in their stories for as long as they have been engaged in reporting, data journalism is more than traditional journalism with more data. Decades after early pioneers successfully applied computer-assisted reporting and social science to investigative journalism, journalists are creating news apps and interactive features that help people understand data, explore it, and act upon the insights derived from it. New business models are emerging in which data is a raw material for profit, impact, and insight, co-created with an audience that was formerly reduced to passive consumption. Journalists around the world are grappling with the excitement and the challenge of telling compelling stories by harnessing the vast quantity of data that our increasingly networked lives, devices, businesses, and governments produce every day.”

“That work is data journalism, or gathering, cleaning, organizing, analyzing, visualizing, and publishing data to support the creation of acts of journalism. A more succinct definition might be simply the application of data science to journalism, where data science is defined as the study of the extraction of knowledge from data.”

“In its most elemental forms, data journalism combines: 1) the treatment of data as a source to be gathered and validated, 2) the application of statistics to interrogate it, 3) and visualizations to present it, as in a comparison of batting averages or stock prices. Some proponents of open data journalism hold that there should be four components, where data journalists archive and publish the underlying raw data behind their investigations, along with the methodology and code used in the analyses that led to their published conclusions.<sup>vii</sup> In a broad sense, data journalism is telling stories with numbers, or finding stories in them. It’s treating data as a source to complement human witnesses, officials, and experts. Many different kinds of journalists use data to augment their reporting, even if they may not define themselves or their work in this way.”

“Data journalism is the practice of finding stories in numbers and using numbers to tell stories,” said Meredith Broussard, an assistant professor of journalism at Temple University.

“Today, the context and scope of data-driven journalism have expanded considerably from its evolutionary antecedent, following the explosion of data generated in and about nearly every aspect of society, from government, to industry, to research, to social media. Data journalists can now use free, powerful online tools and open source software to rapidly collect, clean, and publish data in interactive features, mobile apps, and maps. As data journalists grow in skill and craft, they move from using basic statistics in their reporting to working in spreadsheets, to more complex data analysis and visualization, finally arriving at computational journalism, the command line, and programming. The most advanced practitioners are able to capitalize on algorithms and vast computing power to deliver new forms of reporting and analysis, from document mining applied to find misconduct,<sup>xxvii</sup> to reverse engineering political campaigns,<sup>xxviii</sup> price discrimination, executive stock trading plans, and autocompletions. Data journalists are in demand today throughout the news industry and beyond. They can get scoops, draw large audiences, and augment the work of other journalists in a media organization or other collaboration. By automating common reporting tasks, for instance, or creating custom alerts, one data journalist can increase the capacity of the people with whom she works, building out databases that may be

used for future reporting. “On every desk in the newsroom, reporters are starting to understand that if you don’t know how to understand and manipulate data, someone who can will be faster than you,” said Scott Klein, a managing editor at ProPublica. He continued: The past decade has seen the most dynamic development in data journalism, driven by rapid technological changes. Ten years ago, “data journalism was mostly seen as doing analyses for stories,” said Chase Davis, an assistant editor on the Interactive News Desk at the New York Times. He explained: Great stories, for sure, but interactives and data visualizations were more rare. Now, data journalism is much more of a big tent speciality. Data journalists report and write, craft interactives and visualizations, develop storytelling platforms, run predictive models, build open source software, and much, much more. The pace has really picked up, which is why selfteaching is so important.”

Lewis & Usher (2013) Data journalism, data-driven journalism

Radchenko (2014) Data journalism, database journalism, analytical journalism

“The meaning of the term data journalism seems to be generally clear, while when it comes to the details, it turns out that its application is actually defined very vaguely [23]. Even though there are numerous attempts of reflection, there is still no common understanding of what exactly it means. There are though several terms, sometimes with overlapping meanings, which are aimed to reflect certain subdirections of this trend. Here are some of them. Data driven journalism [24] is normally about producing a digital story which relies heavily on big data sets shown through the lens of interactive visualization. Database journalism [25], in its turn, is first and foremost about creating data sets and databases, which could be further used by data driven journalism. There is also computational journalism [26] that applies computational techniques and the elements of machine/statistical learning to creating a data driven story. A very loose term analytic journalism [27] might also be used in the relation to data journalism as it is sometimes based on data analysis and results in a story describing the conclusions and findings.”

Segel & Heer (2010) data stories

Data visualization is regularly promoted for its ability to reveal stories within data, yet these “data stories” differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have

increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function in place of a written story.

Weber & Rall (2012)	Data stories, computational journalism	<p>“When interactive information graphics are based on large data sets, they are also called data visualization or sometimes “data stories”, a new buzzword that links together two different disciplines: computer science and journalism. This new relationship known as “data-driven journalism” can be seen as a part of computational journalism (Anderson 2011).”</p>
Lorenz (2010)	data-driven journalism	<p>“Data journalists describe the journalistic process as a workflow that consists of: digging deep into data, analyzing and filtering the found data, visualizing it, and forming a story (Bradshaw 2010, Lorenz 2010).”</p>
Tandoc et al. (2015)	data-driven journalism	<p>“We argue that data-driven journalism ought to be distinguished from big data itself. Big data refers to a large set of raw data that requires analysis of some sort (Burns 2014; Couldry and Turow 2014). Data-driven journalism, on the other hand, is the process through which journalists make sense of data—some big, some small—to find angles and details for the story (Bell 2012; Parasie 2014).”</p>
Weber & Rall (2012)	Data-driven journalism	<p>“data-driven journalism, which can be seen as a part of computational journalism [1]. Datadriven journalism is based on large sets of data for the purpose of writing news stories. Data journalists describe the journalistic process as a workflow that consists of: digging deep into data, analyzing and filtering the found data, visualizing it, and forming a story [2, 14].</p>
Baack (2013)	Datadriven journalism	<p>The result significantly differs from traditional news reporting; including illustrations, interactive web applications and reading instructions to make the material accessible. This style of news reporting is called datadriven journalism.</p>
Ausserhofer (2015)	Datenjournalismus	<p>Derartige „computerunterstützte Explorationen im Journalismus“ (Gynnild, 2014, S. 3) haben in den vergangenen Jahren stark zugenommen. Der experimentelle Einsatz von Algorithmen, Daten und sozialwissenschaftlichen Methoden in der Nachrichtenproduktion – von der Recherche über die Datenanalyse bis hin zur Visualisierung – wird im deutschsprachigen Raum vor allem unter dem Begriff Datenjournalismus diskutiert (für eine Begriffstypologie</p>

siehe Coddington, 2014).

Stavelin (2013) Datenjournalismus/  
data driven journalism

“Computational journalism is an emergent field, with high expectations and uncertain boundaries.”

“Computational journalism can presumably be understood as all of these things. As a proposed intersection of journalism and computer science it is a part of information science by both containing a social component (human actors/organizations/social structures) and a technological component of creating and using technological tools. Therefore, theoretical considerations should involve theories that incorporate both aspects.”

““Computer-assisted news reporting refers to anything that uses computers to aid in the news-gathering process” states Melisma Cox in the opening lines of her paper The development of computer-assisted reporting (Cox 2000). The name computer-assisted journalism is also sometimes used, but CAR, short for computer-assisted reporting, is used most often.”

“Data journalism In this context the word data describes digital structured or unstructured raw material that journalists use to investigate, argue, and explain facts.”

““Data journalism’ or ‘computer-assisted reporting’? [...] These are just two terms for the latest trend, a field combining spreadsheets, graphics, data analysis and the biggest news stories to dominate reporting in the last two years. (Rogers 2013)”

“Paul Bradshaw of Birmingham City University explains in The data journalism handbook that the difference between data journalism and “the rest of journalism” is perhaps the possibility to combine the traditional “nose for the news” with large amounts of digital data. “And those possibilities can come at any stage of the journalist’s process: using programming to automate the process of gathering and combining information from local government, police, and other civic sources, as Adrian Holovaty did with ChicagoCrime and then EveryBlock” (Bradshaw in Gray, Chambers, and Bounegru 2012, 2).”

“Analyzing a database or utilizing one for research are activities that are already claimed as precision journalism, data journalism, and CAR. What Holovaty suggests (“Newspapers need to stop the story-centric

worldview” (Holovaty 2006)), and later does with EveryBlock, is to turn online news sites into more granular databases and produce structured information that can be reused at a granular level. An online news story should not be a “blob” or a “text”, but a combination of the elements the story consists of (persons, places, events, dates, etc.) also on the database level, so that the individual pieces can be recombined for multiple and/or future-use contexts.”

“Data-driven investigative journalism: using data to uncover facts - Data storytelling (text, visualisation, video...) - Data-driven applications (mobile or web): serving data to your public - Data journalism website or section”

“In the project’s explanation of what data-driven journalism is they quote Jonathan Stray: “Data journalism is obtaining, reporting on, curating and publishing data in the public interest””

“For now though, we define computational journalism as the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism. In some ways computational journalism builds on two familiar approaches, computer-assisted reporting (CAR) and the use of social science tools in journalism championed by Phil Meyer in Precision Journalism: A Reporter’s Introduction to Social Science Methods (Rowman and Littlefield, 2002). Like these models, computational journalism aims to enable reporters to explore increasingly large amounts of structured and unstructured information as they search for stories (Hamilton and Turner 2009, 4).”

“I define Computational Journalism as the application of computing to the activities of journalism including information gathering, organization and sensemaking, communication and presentation, and dissemination and public interaction with news information, all while upholding values of journalism such as balance, accuracy, and objectivity (Nicholas Diakopoulos 2010, 1).”

“Are precision journalism, CAR, data journalism, database journalism, data-driven journalism, and computational journalism just different names for the same thing? They all have in common a computer-oriented approach to journalism and the

branding of this activity; they all also separate the practitioners from “regular” journalists. They all require specialized skills in more advanced use of computers. To argue that these things are the same, rebranded every few years in order to stay new, fresh, and interesting is not totally wrong. Philip Meyer, one of the men accredited as a pioneer of CAR, argued over 10 years ago that we should stop using the term CAR, as working with computers “no longer defines us”, and that we needed to “move on to a fresher, more ambitious concept” (Meyer in Poynter Institute 1999, 5). Staying fresh is one reason for the plethora of names for this concept. But there are differences. In essence, precision journalism emphasizes the use of scientific methods, CAR emphasizes digital tool use, database journalism emphasizes structure of information storage and retrieval, data and data-driven journalism emphasizes finding stories in data sets, while computational journalism emphasizes the merging of computing and journalistic values in tool creation and method application. There are subtle differences in the semantics, as well as the journalistic foci.”

“Table 1 Comparing software-oriented modes of news production. Precision Journalism Focus Make journalism scientific CAR Data Journalism Utilizing computer tools to produce journalism Distinctive skills Social science methods Advanced computer tool-use Finding, analyzing and presenting data as/in journalism Data wrangling, data storytelling Database Journalism Adding and exploiting the advantages of structure in data journalism Database theory & practice Data-driven Journalism Pursue unknown or presumed stories by following the “data trail” Analytical, investigative research Computational journalism Creating, adapting or using computational tools and method in/as journalism Computational thinking, programming”

“Figure 1: Computational journalism positioned with other types of computer-supported journalistic efforts. The rings bear solid borders in this illustration, but the borders between the practices are actually quite fuzzy. The amount of overlap between the different journalistic types is also made for illustrative purposes.”

Weihnacht & Spiller (2014) Datenjournalismus/d ata driven journalism

Datenjournalismus (data driven journalism) kann – grundlegend als „journalistische Tätigkeit“ (Scholl 1997) gedacht – als Nutzung der Möglichkeiten digitaler Technik im Journalismus verstanden werden.

Datenjournalismus steht somit in der Tradition des Computer Assisted Reporting CAR (I 12, 18, 24, 25, 28, 29, 30), des Precision Journalism (vgl. Meyer 2002; Leßmöllmann 2012), der methodischen Recherche nach Haller (2008) sowie des Knowledge Based Journalism (Patterson 2013) und wird häufig von Wissenschaftsjournalisten oder als Investigativjournalismus betrieben (I 5, 19, 24, 25, 28).

Nach Einschätzung nahezu aller Interviewpartner gibt es noch keine weitreichend akzeptierte Definition. Selbst im „Data Journalism Handbook“ findet sich neben dem Hinweis, dass sowohl das Wort „Daten“ wie auch das Wort „Journalismus“ vieldeutig sei, lediglich eine Ansammlung von Beispielen (vgl. Bradshaw 2012).

Der Kern des Datenjournalismus besteht aus der Sammlung, Analyse und Aufbereitung von digitalisierten Informationen mit dem Ziel einer journalistischen Veröffentlichung. Also handelt es sich 1) um eine spezielle Form der Recherche, die Geschichten aus Datensätzen lesen will; 2) eine spezielle Form der Interpretation von Rechercheergebnissen, die sich an statistischen Maßzahlen orientiert; und allzu häufig auch 3) um eine spezielle Darstellungsform, die Kernbotschaften grafisch und insbesondere als interaktive Webanwendung anschaulich machen will. Vereinzelt wird zusätzlich 4) die Veröffentlichung von Datenherkunft und Rohdatensatz im Sinne des Open Data Ansatzes als elementarer Bestandteil des Datenjournalismus genannt.

De Maeyer  
(2015)

“The definitions of data journalism themselves prove to be slippery. There is a sharp tension between each part of the doublet, data and journalism. The emphasis on data underlines specific challenges and needs, whereas the emphasis on journalism sees the idea as yet another trendy tool for doing good journalism that does not necessarily require particular attention, specific training, or strong organizational policy.”

Dick (2014)  
Hullmann et al.  
(2015)

Information visualization techniques are frequently used to make the data accessible to large public audiences.



## 6.2 Literature Corpus

- (27) Aitamurto, T., Sirkkunen, E., & Lehtonen, P. (2011). *Trends in Data Journalism*. Retrieved from [http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B\\_Hyperlocal\\_Trends\\_In%20Data\\_Journalism.pdf](http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf)
- (09) Appelgren, E., & Nygren, G. (2012). Data Journalism in Sweden – Opportunities and Challenges: A Case Study of Brottspejl at Sveriges Television (SVT). In C. Mills, M. Pidd, & E. Ward (Eds.), *Proceedings of the Digital Humanities Congress 2012*. Sheffield: HRI Online Publications. Retrieved from <http://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren>
- (08) Appelgren, E., & Nygren, G. (2014). Data Journalism in Sweden: Introducing New Methods and Genres of Journalism into “Old” Organizations. *Digital Journalism*, 2(3), 394–405. <http://doi.org/10.1080/21670811.2014.884344>
- (14) Ausserhofer, J. (2015). „Die Methode liegt im Code“: Routinen und digitale Methoden im Datenjournalismus. In A. Maireder, J. Ausserhofer, C. Schumann, & M. Taddicken (Eds.), *Digitale Methoden in der Kommunikationswissenschaft* (Vol. 2, pp. 87–111). Berlin: digitalcommunicationresearch.de. Retrieved from <http://nbn-resolving.de/http://dx.doi.org/10.17174/dcr.v2.5>
- (01) Baack, S. (2013). A New Style of News Reporting: Wikileaks and Data-driven Journalism. In B. Rambatan & J. Johanssen (Eds.), *Cyborg Subjects: Discourses on Digital Culture* (pp. 113–122). CreateSpace Independent Publishing. Retrieved from <http://www.ssoar.info/ssoar/handle/document/40025>
- (20) Bakker, P. (2014). Mr. Gates Returns: Curation, Community Management and Other New Roles for Journalists. *Journalism Studies*, 15(5), 596–606. <http://doi.org/10.1080/1461670X.2014.901783>
- (34) Cohen S., Hamilton JT. and Turner F. (2011). 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.
- (25) Daniel, A., & Flew, T. (2010). The Guardian Reportage of the UK MP Expenses Scandal: A Case Study of Computational Journalism. In *Record of the Communications Policy and Research Forum 2010* (pp. 186–194). Sydney: Network Insight. Retrieved from <http://eprints.qut.edu.au/39358/>
- (37) Davenport, L., Fico, F., & Detwiler, M. (2000). Computer–assisted reporting in Michigan daily newspapers: More than a decade of adoption. Presented at the Association for Education in Journalism and Mass Communication (AEJMC) National Convention, Phoenix, Arizona.
- (36) Davenport, L., Fico, F., & Weinstock, D. (1996). Computers in newsrooms of Michigan’s newspapers. *Newspaper Research Journal*, 17(3-4), 14–28.
- (28) De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F., & Le Cam, F. (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. <http://doi.org/10.1080/21670811.2014.976415>
- (17) Dick, M. (2014). Interactive Infographics and News Values. *Digital Journalism*, 2(4), 490–506. <http://doi.org/10.1080/21670811.2013.841368>
- (15) Felle, T. (2015). Digital Watchdogs? Data Reporting and the News Media’s Traditional “Fourth Estate” Function. *Journalism*, 1464884915593246. <http://doi.org/10.1177/1464884915593246>
- (11) Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States: Beyond the “Usual Suspects.” *Journalism Studies*, 16(4), 467–481. <http://doi.org/10.1080/1461670X.2014.939852>
- (21) Garrison, B. (1999). Newspaper Size as a Factor in Use of Computer-Assisted Reporting. *Newspaper Research*

*Journal*, 20(3). Retrieved from <http://com.miami.edu/car/baltimore1.htm>

- (18) Gynnild, A. (2014). Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets. *Journalism*, 15(6), 713–730. <http://doi.org/10.1177/1464884913486393>
- (04) Hannaford, L. (2015). Computational Journalism in the UK Newsroom: Hybrids or Specialists? *Journalism Education*, 4(1), 6–21. <http://doi.org/http://hdl.handle.net/2173/575953>
- (24) Howard, A. B. (2014). *The Art and Science of Data-Driven Journalism* (pp. 1–144). New York: Tow Center for Digital Journalism. Retrieved from <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>
- (05) Hullman, J., Diakopoulos, N., Momeni, E., & Adar, E. (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* (pp. 1170–1175). New York: ACM. <http://doi.org/10.1145/2675133.2675207>
- (03) Karlsen, J., & Stavelin, E. (2014). Computational Journalism in Norwegian Newsrooms. *Journalism Practice*, 8(1), 34–48. <http://doi.org/10.1080/17512786.2013.813190>
- (10) Knight, M. (2015). Data Journalism in the UK: A Preliminary Analysis of Form and Content. *Journal of Media Practice*, 16(1), 55–72. <http://doi.org/10.1080/14682753.2015.1015801>
- (22) Lewis, S. C., & Usher, N. (2013). Open Source and Journalism: Toward New Frameworks for Imagining News Innovation. *Media, Culture and Society*, 35(5), 602–619. <http://doi.org/10.1177/0163443713485494>
- (02) Lewis, S. C., & 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. <http://doi.org/10.1080/21670811.2014.895504>
- (23) Lugo-Ocando, J., & Brandão, R. F. (2015). Stabbing News: Articulating Crime Statistics in the Newsroom. *Journalism Practice*, 9, 1–15. <http://doi.org/10.1080/17512786.2015.1058179>
- (07) Parasie, S. (2015). Data-driven Revelation? Epistemological Tensions in Investigative Journalism in the Age of “Big Data.” *Digital Journalism*, 3(3), 364–380. <http://doi.org/10.1080/21670811.2014.976408>
- (06) Parasie, S., & 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. <http://doi.org/http://dx.doi.org/10.1177/1461444812463345>
- (26) Radchenko, I., & Sakoyan, A. (2014). The View on Open Data and Data Journalism: Cases, Educational Resources and Current Trends. In D. I. Ignatov, M. Y. Khachay, A. Panchenko, N. Konstantinova, & R. E. Yavorsky (Eds.), *Analysis of Images, Social Networks and Texts* (pp. 47–54). Cham: Springer International Publishing. Retrieved from [http://link.springer.com/chapter/10.1007/978-3-319-12580-0\\_4](http://link.springer.com/chapter/10.1007/978-3-319-12580-0_4)
- (33) Royal C (2012) The journalist as programmer: A case study of the New York Times Interactive News Technology Department. #ISOJ *The Official Research Journal of the International Symposium on Online Journalism* 2(1).
- (35) Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Trans. Visualization and Computer Graphics*, 16(6), 1139–1148. <http://doi.org/10.1109/TVCG.2010.179>
- (38) Smit, G., Haan, Y. de, & Buijs, L. (2014). Visualizing news: Make it work. *Digital Journalism*, 2(3), 344–354. <http://doi.org/10.1080/21670811.2014.897847>
- (31) Stavelin E. (2013). Computational journalism: When journalism meets programming. Dissertation, Bergen: University of Bergen.
- (40) Tabary, C., Provost, A.-M., & Trottier, A. (2016). Data journalism’s actors, practices and skills: A case study from Quebec. *Journalism*, 17(1), 66–84.

- (32) Tandoc Jr. EC. and Oh S-K. (2015.) Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism Studies*. DOI: 10.1080/1461670X.2015.1104260
- (19) Uskali, T., & Kuutti, H. (2015). Models and Streams of Data Journalism. *The Journal of Media Innovations*, 2(1), 77. <http://doi.org/10.5617/jmi.v2i1.882>
- (12) Weber, W., & Rall, H. (2012). Data Visualization in Online Journalism and Its Implications for the Production Process. In *2012 16th International Conference on Information Visualisation (IV)* (pp. 349–356). Montpellier: IEEE. <http://doi.org/10.1109/IV.2012.65>
- (30) Weber, W., & Rall, H. (2013). "We Are Journalists": Production Practices, Attitudes and a Case Study of the New York Times Newsroom. In W. Weber, M. Burmester, & R. Tille (Eds.), *Interaktive Infografiken* (pp. 161–172). Berlin: Springer. Retrieved from [http://link.springer.com/chapter/10.1007/978-3-642-15453-9\\_9](http://link.springer.com/chapter/10.1007/978-3-642-15453-9_9)
- (13) Weihnacht, S., & Spiller, R. (2014). Datenjournalismus in Deutschland: Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik*, 59(4), 411–433. <http://doi.org/10.1007/s11616-014-0213-5>
- (16) Young, M. L., & 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. <http://doi.org/10.1080/21670811.2014.976409>
- (39) Zanchelli, M., & Crucianelli, S. (2012). *Integrating data journalism into newsrooms*. International Center for Journalists. Retrieved from [http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english\\_0.pdf](http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english_0.pdf)