mediální studia

media studies

Turning Data into News:
Five Different Skill Sets of Journalists Working with Data

Liis Auväärt & Ragne Kõuts-Klemm

To cite this article:
Auväärt, L. - Kõuts-Klemm, R. (2025). Turning Data into News: Five Different Skill Sets of Journalists Working with Data. *Mediální studia*, 19(1), 6-30.

ISSN 2464-4846
Journal website: https://www.medialnistudia.fsv.cuni.cz/

JOURNAL FOR CRITICAL MEDIA INQUIRY

1/2025

TURNING DATA INTO NEWS: FIVE DIFFERENT SKILL SETS OF JOURNALISTS WORKING WITH DATA

LIIS AUVÄÄRT & RAGNE KÕUTS-KLEMM
University of Tartu, Estonia

ABSTRACT
Since the 2010s, but even more so in the post-COVID era, numerous studies discuss new ways to do journalism using data. Yet little emphasis has been put on the crucial question: what skills are needed to work with different types of data? – the answer to which could assist journalism educators and media houses alike. Based on a content analysis of scientific research articles from 2008–2023 (n=51) we propose five profiles of data journalists based on their data skills: the specialised data journalist, the data analyst, the techie of a newsroom, the daily editor and the special beat editor. We argue, that looking at the issue of various data in journalism following a skill-based logic will contribute to updated journalism curriculums and more precise job requirements by the newsrooms.

KEYWORDS
data • journalism • journalistic skills • numeracy • statistics • Big Data

INTRODUCTION

The field of data journalism research has gained noticeable momentum since the 2010s, with numerous studies discussing new ways to do journalism using data (Ausserhofer et al., 2020; Beiler et al., 2020, Erkmen, 2024). Yet, as derived from a recent systematic review on data journalism, little emphasis has been put on what data skills journalists actually possess (Erkmen, 2024). The main focus of studies has been on empirical evidence of how newsrooms integrate data journalism into their work (Erkmen, 2024, p. 63). The other popular research topic has focused on the outcomes of data journalism – e.g., how data are used to tell the stories (Auväärt, 2023; Chaparro-Domínguez & Díaz-Campo, 2021; Loosen et al., 2020; Ojo and Heravi, 2018; Young et al., 2018) or how the stories are made understandable for users through visualisation and data presentation techniques (Engebretsen et al., 2018). The latest studies take the holistic view and discuss the epistemologies related to data (Morini, 2023; Ramsälv et al., 2023) or point to the innovation data journalism can bring with for media and societies (Wu, 2024).

The study at hand will attempt to contribute to the more systematized and complex understanding of skills needed by journalists to do different types of data journalistic work. The necessity for such an approach is backed in several ways by the ongoing development of data journalism. Firstly, the need to acquire skills to respond to datafication has only become more pressing in newsrooms (especially due to the years of data-heavy COVID-19 coverage worldwide). In light of the open data movement, data can be viewed as a prerequisite for generating knowledge and journalists as "data intermediaries", who use their skills to refine it and thus create knowledge for the public (Baack, 2015). "Data-base analysis" has been argued superior over "observational expertise" in a variety of fields such as sports and finance (Kallinikos, 2009), while "observational expertise" is arguably as a traditional journalistic skill and the need for data-base analysis a result of changing times (Thurman et al., 2017). Hence although previous years saw a noticeable shift to downsizing newsrooms, shrinking budgets and advertising revenues, "data journalism has come into the spotlight as one of the few expanding areas in newsrooms" (Beiler et al., 2020, p. 1571).

Secondly, the relevance of data for journalistic work is highlighted by the diversification of journalism occupations. Some examples of this being "programmer/journalist", "journalist/developer", "hacker/journalist" (Royal, 2012), "news automation specialist" (Beiler et al., 2020) and "programmer-journalist" (Lewis & Usher, 2014; Parasie and Dagiral, 2013) in newsrooms. A popular distinction exists between "newshound" and "techie": the former embodying a traditional journalistic way of handling and engaging with data and the latter an emergent journalistic approach to data based on more computational logics and mindsets (Borges-Rey, 2020). The best data journalism stories tend to involve a wide variety of data, and bringing it together takes a lot of time, diverse teams and excellent data skills, as the analysis of award-winning data journalism shows (Loosen, 2020). Yet in "day-to-day" data journalism (Zamith, 2019), there is a strong reliance on data provided by official institutions and/or other non-commercial organisations (NGOs, research institutes, etc.), which is also accompanied by the agenda of the data providers (Beiler et al., 2020; Loosen et al., 2015; Loosen et al., 2020; Parasie & Dagiral, 2013; Van Witsen, 2020, Zamith, 2019).

While data is a complex phenomenon that can be extracted from physical and social reality based on various principles (e.g., inductive and deductive, qualitative and quantitative approach) and with different scientific tools (from computational sciences to social sciences and humanities), the multiple options to use data by journalists deserve attention. We argue that data usage for reporting is determined and limited by the character of the data – interpretation of sociological survey data requires different knowledge than usage of the data from weather forecasts or sports results. However, only few examples of more detailed data definitions exist in data journalism research. As an example, Steensberg (2021) researched the "quantitative claims" in news stories, but did not specify the type of data. He defined "quantitative claims" (Steensberg, 2021) as verbalised expressions about something being more or less, with references to numbers. Knight (2015) analysed "data elements" in UK newspapers and defined them as: numbers in text, timelines, static maps, dynamic maps, graphs, infographics, tables, figures, lists of numbers and numerical pull quotes (Knight, 2015, pg. 61). Moreover, the triumph of Big Data created numerous new possibilities for data journalism (Portilla, 2018; Ramsälv, 2023; Veglis & Bratsas, 2017), but also: the use of Big Data requires more advanced data processing and interpretation skills (Hammond, 2018), introducing the methods of data science in the newsrooms (Hermida & Young, 2019). Already in 2016, Stencel and Perry surveyed 31 news organizations' hiring priorities, dividing core skills into foundational skills (e.g. editing and writing) and transformational skills (needed to adapt to ongoing changes in "news audience, distribution, editorial practices and presentation"). The top five skills listed by the organizations were transformational, with the top skill (a priority for 71% of news organizations) being "coding/development"; "user data and metrics" was near the top.

Research on human information processing suggests that numbers and statistics can have authority as facts about reality in a modern world (Porter, 1996; van der Bles et al., 2020). Mastering data can therefore be an opportunity to provide information that is trusted. We will take this claim a step further by inviting the data journalism enthusiasts to think more deeply about the characteristics of data and the skills needed to manage it. The two following subsections thus ask: 1) what are ***data***? and 2) which skills are needed for data-related journalism?

1. WHAT ARE *DATA*?

Following the classifications in social sciences, one can distinguish between data in terms of form (qualitative or quantitative), structure (structured, semi-structured or unstructured), source (captured, derived, exhaustive or transient), producer (primary, secondary or tertiary) and type (indexical, attribute or metadata) (Kitchin, 2014, p. 4). Digitalization and the need to analyze digital human 'traces' also bring the computational sciences to the spotlight to develop a multidisciplinary data-centered approach. (Asamoah et al., 2015, Parti and Szigeti, 2021). Due of digital technology the human ability to get significant/interpretative units of information from their environment has been rising (Rowe, 2023) and thus the need to develop critical data literacy. However, current research shows the prevalence of naïve data optimism and uncritical approach towards datasets, dealing with them as passive and value-free elements (Bhaskaran et al., 2024).

Among data processors there are several specializations needed. Professionals who work with data have to combine expertise in computing, statistics, experiment design, interpretation and analytics with fundamental business knowledge and acumen in order to pose the right questions (Hopkins et al., 2010). Data science is multidisciplinary and include subjects of computer science, mathematics, and statistics, including specific competencies in data security, data ethics, data governance, data integration, and data visualization (Coners et al., 2025). Research shows that the work of investigative data journalists significantly resembles data science work practices (Showkat & Baumer, 2021).

As suggested by Fotopoulou (2020, p. 2): "making sense and meaning of data and big datasets, such as electoral data or health data, is not only a technical but also a sociocultural process," thus creating a need for data literacies that move beyond enhancing quantitative analysis and technical skills. In trying to bring more clarity to data journalism research, one could start with classifying data based on the type of data collection, as the purpose of data collection is highly relevant to understanding the usability of and possible biases in data (boyd & Crawford, 2012; Gitelman, 2013). Following the historical development of social data use in social sciences (Raftery, 2001), one can classify the data as a) measured facts from the "physical" world and routinely given numbers derived from the real-life events – i.e. statistics, b) sociological and research data collected using carefully designed, but clearly limited methods, and c) data that are collected "by themselves", i.e. Big Data derived from the use of digital technology, including the internet.

Starting with the first, research shows that societal issues, such as census results, crime reports, health and science issues, and business and economics are common themes covered by journalists (Cushion et al., 2016; Loosen et al., 2020). This type of data are given in the form of particular numbers: economic growth, GDP, number of immigrants, weather forecasts etc. Numerical data are quantitative and structured. They can be derived from different sources and different producers and can be of different types in the classification described by Kitchin (2014). Journalism has developed routinised forms to present this kind of data: e.g., financial news constructed by algorithms following a pre-set format, algorithms for data mining etc (Diakopoulos, 2019; Miroshnichenko, 2018; Sirén-Heikel et al., 2023). Research shows complex ways in which automatic discovery can find newsworthy themes from sequenced data (e.g., ***k***-Sketch query by Fan et al., 2017), which a human journalist can then work with, using journalist's specialist skills: interviewing, critical thinking and understanding newsworthiness (Ferrucci, 2018).

The second type of data that journalists encounter daily is pre-processed and interpreted by third parties (e.g., crime reports, health and science issues). These are collected by governments, public institutions or NGOs, or purchased by research institutions. These data can be in the form of different studies for specific research purposes or statistics collected by officials with the aim of better administration (Porter, 1995) and governance. Stalph (2017) has shown that "the everyday data journalism" depends mainly on "the pre-processed data drawn from domestic governmental bodies" (2018, p. 1332). Ongoing criticism states that journalists do not act as watchdogs in society because they replicate the official interpretation of events and processes or even empower the voices of those who already have power to determine discourses (Lugo-Ocando & Lawson, 2017).

The abandoning of the gatekeeper role by journalism in the case of data offered by official bodies is usually connected with the lack of the necessary skills (Brechman et al., 2009; Van Witsen, 2020) and work processes that limit the time for in-depth investigations (Reich & Godler, 2014). The lack of knowledge of data collection methods causes the spread of biased data interpretations (Garz, 2014).

Thirdly, journalists deal with Big Data, which are often in textual or audio-visual form. In both cases the "volume" of data makes it hard to handle without specialized data skills, yet the possibilities for journalistic projects fuelled by Big Data technologies are vast and rich: ranging from consumer choice to geographical position, web movement and behavioural information (Bolin & Andersson Schwarz, 2015). The availability of non-pre-interpreted data has been growing since the intervention of Big Data in economics and governance, and has been accompanied by the open data movement

(Baack, 2015). With the open data movement, more databases are open access: both a boon and misfortune for journalists. Rowe (2023, p. 4) touches on this saying: "Handling Big Data requires a wider and new digital skills set, largely based on machine learning, artificial intelligence and coding, in addition to greater knowledge of computing technology".

Finally, sociological data are sometimes and in limited capacity collected by the newsrooms themselves (e.g., polls and questionnaires), but the quality of such datasets and how they are used needs further academic research.

2. WHICH SKILLS ARE NEEDED FOR DATA-RELATED JOURNALISM?

Skills are an important concept in occupational studies and economics (Green, 2011) because they can explain differences between individuals in the performance of different tasks. For example, numeracy has been shown to be one of the most powerful predictors of gender wage gap (Battisti et al., 2023; Hanushek et al, 2015). We hereby also propose to look at the issue of various data in journalism following a skill-based logic.

According to Fischer's (1980) cognitive skill theory, "skills are ordered in levels of growing complexity, with a specific skill at one level built directly from specific skills at the preceding level, and a set on transformation rules that relate these levels to each other" (Fischer, 1980, p. 477). Skill studies conceptualise the hierarchical structure of skills, where the skills are organised similar to three with increasing sophistication (Fisher, 1980; Sharma, 2017). However, they can be analytically divided into foundational and specialist skills, as shown by Wolff et al. (2016). These foundational skills (e.g., general knowledge how to process data) can be used at different levels as specialist skills, such as the ability to convert data or to carry out a reliability analysis.

Data literacy is a cognitive skill, similar to mathematical proficiency, numeracy and statistical literacy (Kilpatrick, 2001; Sharma, 2017, Vahey et al., 2012). Cognitive skills have been empirically shown to be acquired over the life course (Lechner et al., 2021) and to be convergent, meaning that some skills underpin the performance of others (Lee et al., 2012). Skill theory suggests that skills require a compelling environment to develop, i.e. through increasing experience of operating in a complex environment, a person can develop their skills to the highest level they are capable of (Doppelt, 2019; Fischer, 1980; Lechner et al., 2021). In the context of journalism, this means that the more challenging the newsroom is to update skills, the more journalists will develop their skill profiles to the highest level required to do their jobs effectively.

Data journalism can be described as being closely tied to maths and technological expertise, yet with an emphasis on the classical "nose for news" (Royal, 2012, p. 22). Appelgren & Nygren (2014) mark, that when creating data journalistic projects, the journalistic angle comes first, not data or technology. Almost a decade ago it was noted in an analysis of the Chicago Tribune that data journalists found most stories from already released data, thus making it the reporter's job to identify paths in this data (Parasie & Dagiral, 2013). Loosen et al. (2020), analysing international data journalism competition, noted this as a continuing and growing trend. Hence, one possible way to study data literacy through the workflow of journalists is to implement a five-step data-processing model (Kõuts-Klemm, 2019): increasing the journalist's ability 1) to find data, 2) to evaluate data quality, 3) to interpret data in context, 4) to present data through journalistic means, and 5) to consider the needs and capacities of reception of audiences. Returning to Wolff et al. (2016) the first three steps of this model would be foundational skills and the two remaining steps – specialist skills characteristic of journalistic work.

Concentrating on data literacy is key here – being overly perceived as „technical abilities like extracting data, making statistical analysis, creating visualizations, interpreting and reporting appropriately" (Raffaghelli & Stewart, 2020). Researchers agree almost by consensus, that data journalism requires certain computational skills, which are often compensated for by working in multi-skilled teams. While "ordinary" (De Maeyer et al., 2015) or "day-to-day" (Zamith 2019) data journalism "is manageable by one individual [and] can be done on a daily basis" (De Maeyer et al., 2015, p. 410–411), "thorough" (De Maeyer et al. 2015) data journalism relies on a team with skill sets complementing one another. For example, Borges-Rey (2020), researching the devolved nations of the United Kingdom, describes journalists teaming up with graphic designers and coders to compensate for "the absence of certain advanced computational skills and/or the restricted access to certain information" (Borges-Rey, 2020, p. 926). In environments such as North America a trend has been observed to specialize in two profiles: one responsible for collecting and analysing data, the other creating visualizations and apps (Bisiani et al. 2023, p. 3).

Yet most often in journalism research data journalists are discussed with no distinction between different roles and skills. Also, newsrooms often invest in a solitary data journalism hire, who works on full-time data production, but does not have a designated beat (Boyles & Meyer, 2017). This leads to modern data journalist being described as self-taught, resourceful and multi-skilled (Appelgren, 2017; Beckett & Deuze, 2016; Bisiani et al., 2023; Himma-Kadakas & Palmiste, 2019; Larrondo et al., 2016; Örnebring, 2010; Örnebring & Mellado, 2018).

We argue that a ***skill-based differentiation*** of data journalists – a step further from "ordinary"/ "day-to-day" vs "thorough", looking at specific roles carried out in newsroom – is in better accordance with professional roles in modern newsrooms. Although there are some handbooks and textbooks that thematize the variety of data in journalism (e.g., Gray et al., 2012; Livingston & Voakes, 2005), this study is designed to view how the nature of data can influence the complexity of the skills journalists need to work with it.

Thus, the following questions are posed:

- ***RQ1: What kind of skills related to journalistic information processing are revealed as necessary for a data journalist and how do these skills relate to different types of data?***
- ***RQ2: Are there associations between how data is defined by the academic literature and what journalistic skills are discussed?***

3. METHODOLOGY

The following research is based on an analysis of academic literature, focusing on data skills of journalists and/or journalism students mentioned/ analysed in scientific articles published during last 15 years. The same time frame – 2008–2023 – was used in a systematic literature review by Erkmen (2024), noting that in the year 2008 the term "data journalism" was first used in Web of Science, the oldest international bibliometric database. Although for Erkmen (2024) the earliest manuscript meeting the inclusion criteria was not published before 2013, we decided to keep 2008 as a starting point, as this range would encapsulate a full body of studies published on the topic and the search string of this study could return hits before the year 2013.

The decision to include articles addressing data in terms of journalism education was based on the fact that these articles reflect on the current state of teaching data skills to future journalists and propose changes in curricula to match the hiring needs of the changing media market.

As a scientific tool, the literature review presented here has several purposes. It can be viewed as a way to "examine old theories and propose new ones, consider where the balance of evidence lies in relation to a particular topic" (Petticrew & Roberts, 2006, p. xiii), and offer guidance to researchers planning future studies. Although a fundamental component of academic research, literacy review as a method comes with several potential limitations, such as possible lack of transparency and/or reproducibility if the stages and phases of the study have not been explained with necessary clarity (Kraus et al., 2022) – furthermore, literature reviews may self-sustain biases in the field's boundaries and priorities (Gond et al., 2023). One might claim, that human-collected literature review as a method for knowledge extraction will be even more questionable in the development of artificial intelligence, where AI can substantially contribute to knowledge and theory development (Wagner et al., 2022). However, AI tools still lack crucial abilities (Bolanos et al., 2024); thus, human contribution has value.

Keeping all of the above in mind, we 1) aim to examine the research done so far, 2) group the information presented in the articles and 3) discuss potential applications of this knowledge in future research. We describe the stages and phases of the review in detail, to allow transparency and replicability.

To form a corpus of literature, the following steps were taken. First the decision was made to gather scholarly works from the major online research databases Scopus and EBSCO, as these have become international platforms for scholars to publish research. We focused our research to last 15 years (2008–9/2023), a time frame also used by Erkmen (2024).

In an electronic search strategy relevant key words and appropriate subject headings are important elements (Spry & Mierzwinski-Urban, 2018). To gather the corpus of literature, a preliminary table of search terms was constructed. The table included relevant key words, synonyms or related phrases. The most general key term was "datafication", which can be seen as the "quantification of aspects of life previously experienced in qualitative, non-numeric forms, which are then tabulated, analysed and visualized" (Engebretsen et al., 2018, p. 1). This was followed by the term "journalism" and "data skills" to sharpen the focus. The preliminary search string was complemented by adding new search words found in the titles and abstracts of articles identified as potentially relevant to the study at hand and the searches repeated, until no new hits identified as relevant were added (Gehanno et al. 2009. Using Boolean operators, a search string was finally composed as follows: *(Data*) AND (journalist OR reporter OR (data journalist) OR journalism OR (data journalism)) AND ((data skill*) OR (data competence) OR competence OR skill* OR (professional skill*))* (see Table 1).

Table 1: The search terms (relevant key words, synonyms or related phrases) used to create a search string for online databases. Source: Authors

Search terms	Datafication	Journalism	Data skills
Related key words	Data	Data journalism Journalist Data journalist Reporter	Data literacy Data competence Professional skills Skills

This search was conducted in Scopus, searching article titles, abstracts and keywords, and ended up with 447 results. Additional filters were set to 1) focus on the subject area of social sciences. This was done because after reviewing the total list of articles it was evident, that the vast majority of research related to journalistic skills comes from this field: 250+ articles, followed by the field of medicine (80+ articles) and thirdly computer science (60+ articles). The interest in data journalism also differs by field: computer sciences, for example, concentrate more on potential tools used for data journalism, while social sciences take an interest in which attributes and skills are required to work in newsrooms. As such, the focus on literature from the social sciences seemed appropriate and justified. Additional filters were set for 2) the type of text to articles/conference papers published in English, leaving 180 results, and 3) all key words matching journalism/journalists were selected, leaving 116 results.

A second search was conducted in EBSCO Discovery with the same search string. Applying the filter of key words matching journalism/journalists, the results were compared with those from Scopus. Four additional texts were added to the selection.

The 120 texts were collected from listed databases and abstracts reviewed. Those texts that a) did not somehow reflect upon the data skills of journalists/ journalism students, b) were unavailable as a whole, or c) were not written in English were excluded. As a result, a selection of 51 articles was formed (for reviewing purposes: you can see the list of articles in the appendix. This can be made available to fellow researchers as a request from the author(s), should it be lengthy to add to the article). The number of articles found is in line with a literature review by Erkmen (2024) covering the same period, who found that only less than half of the articles (97 in total) dealt with journalists' perceptions and practices – topics under which skills have been researched.

To answer the research questions, we started by coding the parts of the texts (sentences or sections consisting of several sentences) that 1) indicated some form of data, 2) referred to the journalist doing something with data, or 3) mentioned journalistic skills needed to work with data. The initial coding scheme was created inductively following the previous studies – it is further discussed below – and tested by pilot coding. During the pilot coding sub-categories for the main categories were created (see Table 2).

Both for coding and analysis MaxQda software – MaxQda 2022 – was used. While software-based qualitative analysis has been criticised for its "code-and-retrieve" approach (Bazeley & Jackson, 2013), we followed Mitchell & Schmitz's (2023) suggestions on how not to "reduce complex qualitative data to mere lists of codes and frequency counts" (Mitchell & Schmitz, 2023, p. 167). Four steps were followed: 1) we created a sample of articles that was as coherent as possible (only peer-reviewed articles addressed journalists' data skills), 2) we selected a sample of medium size, i.e. small enough for manual coding, 3) we read all documents carefully before coding and further processing, and 4) we followed the skills theory in interpreting the patterns illuminated by the software.

In detail, the references to data were categorized based on common sociological classifications: 1) directly mentioning data in terms of numbers, 2) sociological research data (gathered with the specific aim and with the help of social scientific methods: polls, surveys, census data etc.), 3) statistics (data gathered regularly for administrative purposes: different state registries etc.), and 4) Big Data (data gathered "by themselves" as a result of the usage of digital technology). The last category was divided into structured (geolocation, weather, financial transactions etc.) and unstructured (textual data from social media, collections of speeches, documents etc.) data. The term was given a code, if it had been put into context by the author(s) of the article being coded. For example, the code "2. data = numbers" was given, if the term was accompanied by clarification like *measured knowledge, expressed in numbers*.

The coding of type of data was relevant for two purposes: 1) to show the variety of treatment and 2) to relate the data definitions to the skills of journalists mentioned in the texts.

The pure mentioning of data was subject of mixed-methods approach – by analysing the mentioning of the word in relation to other words there is possibility to reveal the connotations of the concept, e.g., "data scraping" is activity characteristic to internet data that can be in different forms, whereas "dataset" already means that there is an inner organization of input provided.

The other part of the coding scheme, skills, was based on the previously presented overview of journalism theoretical literature (e.g., Attfield et al., 2009; Bradshaw, 2011; Kõuts-Klemm, 2019) and followed the stages of journalistic work process. The skills were organized in a linear manner: 1) establishing an initial idea, 2) preparations for info gathering, 3) obtaining information, 4) analysing the gathered material, 5) composing articles, and 6) distributing finished work. These categories were coded if a clear connection with data was made in the text. The coding of texts was carried out by both authors. After a test sample of five articles, the coding scheme was discussed between authors and specified even further. Inter-coder reliability was calculated based on three articles and was 82%. Using software for coding meant that the coders marked parts of the text, e.g., a single word, a few words, sentences, or even longer parts of the text, as a code. The scope of a code in the text was also a topic of coding agreements. Later, the software relates not only the codes as such to each other but includes the particular parts of the texts also.

To further analyse the connections between the findings, hierarchical cluster analysis was used (MaxQda, 2022). This technique helps visualize group affiliations as codes are grouped based on similarity. The basis for this calculation is a distance matrix showing how closely connected the assigned codes are in texts and how tightly tied to another. In our case it visualizes which types of data and which journalistic skills were most often related in the papers and how strongly these were linked. For example, if ***sociological data*** is mentioned as being used by a journalist, is ***quantitative literacy*** also discussed in this academic paper.

Relating data and skills in scientific articles enabled to assess the sophistication of the data journalism research: if researchers had created a clear link between different types of data and different skills or not.

4. RESULTS

Since the articles represented a variety of topics related to data (in) journalism, from academic education (e.g., Davies & Cullen, 2016; Kashyap & Bhaskaran, 2020) to overviews of data journalistic practices in different countries/regions (e.g., Fink & Anderson, 2015; N P Lewis & Nashmi, 2019) or discussing journalistic skills (e.g. Rodríguez & Clark, 2021; Thurman et al., 2017) the amount of effort given to defining “data” or “skills” also varied greatly. But because all of the collected articles were derived from academic databases via the same search string, how deeply the base terms “data” and/or “skills” were discussed in the texts helped to illustrate how “obvious” or “self-evident” these concepts seemed to be related to data journalism.

In some articles the term “data” was left vague and tied to journalistic skills only in a paragraph or two, while in others they were discussed as very much linked. Creating an Excel “heat map” revealed that most connections were in articles dealing with teaching data journalism (e.g., Kashyap & Bhaskaran, 2020; Treadwell et al., 2016).

4.1. Modern skill-set: story-telling plus technical skills

The frequent occurrence of the concept of Big Data explains the most prominent category concerning journalistic skills, “establishing initial idea”, which included the attributes passionately discussed in data journalism research, i.e. “technical skills”. Throughout the coding, we differentiated between “basic technical skills” and “advanced technical skills”: the first indicated that a journalist was able to use technical tools to perform data-related journalistic tasks (e.g., spreadsheets), but did not have the know-how to create such tools from scratch. The second indicated that a journalist had to be at least somewhat skilled in coding, programming, creating templates etc. Judging from the sample of texts, "advanced technical skills" were most frequently discussed in addressing journalistic work related to data.

As for traditional skills, "narrating, storifying" was named notably often, as was "nose for news". For example, phrases such as the journalist had to "find and extract stories from big data" (Green, 2018), "produce stories through a combination of software programming and storytelling skills" (Kosterich & Weber, 2019) or that "data journalists must engage their publics through more emotive reporting, without losing sight of its factual strength" (Stalph & Borges-Rey, 2018) were used. These findings emphasize that the need for story-telling skills had not been lost in the wave of datafication. Presenting data in a clear and compact way was a journalistic challenge, echoed by the frequency with which "visualizing" (e.g., creating maps, timelines and other infographics) was mentioned in the sample (see Table 2).

Table 2: Frequency of the codes and categories in the sample of articles. Source: Authors

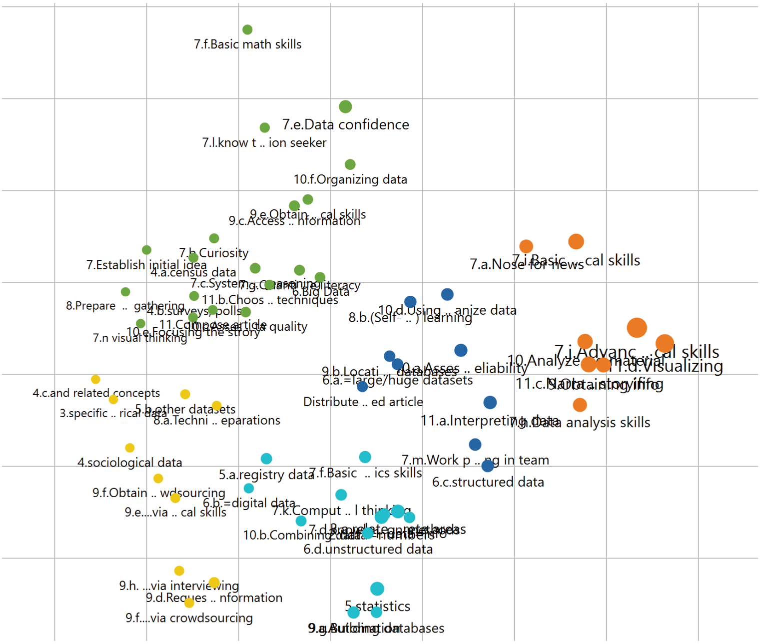
Category/code	Frequency of a code	Frequency of a category
1. data = information	**23**	**23**
2. data = numbers	**45**	**45**
3. Specific numerical data	**3**	**48**
3.a related to concrete areas	45	
4. Sociological data	**4**	**22**
4.a census data	9	
4.b surveys, polling data	8	
4.c related concepts	1	
5. Statistics	**47**	**78**
5.a registry data	22	
5.b other datasets	9	
6. Big Data	**15**	**112**
6.a =large/ huge datasets	26	
6.b =digital data	11	
6.c structured data	34	
6.d unstructured data	26	

7. Establishing initial idea	**4**	**495**
7.a Nose for news	46	
7.b Curiosity	9	
7.c Systematic thinking/reasoning	16	
7.d Knowledge of different data collection methods	30	
7.e Data confidence/literacy	39	
7.f Basic maths skills	10	
7.f. Basic maths and statistical skills	30	
7.g Quantitative literacy	17	
7.h Data analysis	49	
7.i Basic technical skills	65	
7.j Advanced technical skills	108	
7.k Computational thinking	25	
7.l Knowing legal rights as a seeker of information	11	
7.m Work planning in teams	33	
7.n Visual thinking	3	
8. Preparing for info gathering	**2**	**40**
8.a Technical preparations	7	
8.b (Self-directed) learning	31	
9. Obtaining info	**58**	**220**
9.a Automation	31	
9.b Locating information, navigating databases	23	
9.c Accessing information	18	
9.d Requesting information	19	
9.e Obtaining information via technical skills	23	
9.f Obtaining information via crowd-sourcing methods	16	
9.g Building databases	23	
9.h Obtaining information via interviewing	9	
10. Analysing gathered material	**61**	**191**
10.a Assessing data reliability	39	
10.b Combining data	18	
10.c Assessing data quality	15	
10.f Organising data	17	
10.d Organising data using technological tools	34	
10.e Deciding on the final focus of the story	7	
11. Composing articles	**8**	**220**
11.a Interpreting data	42	
11.b Choosing the techniques	12	
11.c Narrating, storifying	66	
11.d Visualising	92	
12. Distributing	**19**	**19**

4.2. Distinguishing data journalists via skill-sets: 5 types

As the next step in the analysis, the co-occurrence of the different codes and categories was observed. Creating a proximity matrix allowed visualize how closely connected these codes were in the texts (see Figure 1). To clarify, clusters are sets of objects that are grouped based on similarities (Keller & Achatz, 2019, p. 425), in our case based on the co-occurrence of the coded phrases in two main categories: 1) characteristics of the data, and 2) journalistic information processing and related skills. If the same codes occur repeatedly in a number of documents together, the clustering method classifies them into the same cluster. In the graph, five distinct clusters emerge.

Figure 1: Proximity matrix: five clusters of articles (occurrence of codes in the same document) (colours of codes). Source: Authors own processing and software.



Type 1: the specialised data journalist. Cluster 1 (marked with the colour green on Figure 1) represents the type of journalist with the widest skill-set for working with data. Such person is data confident and knows the legal rights of an information seeker. Such journalist has basic technical skills to combine and interpret societal data (both sociological data and statistics) created by a third party. Clearly concepts dealing with sociological data are

closely tied to the first stage of the journalistic work process: being curious about a matter and establishing the initial idea. A part of the same process is the ability to think visually and focus the story, which can be summarized as seeing the story behind the (numerical) data. The mid-section of the cluster visualizes the journalist's preparations to access and obtain the data, leading to requests for data (via surveys and polls, using census data, Big Data) and tied to organizing the data. Knowledge of data collection methods help journalists interpret their data independently from the frames presented by data holders. Note, that focusing the story and visual thinking are closely linked. We propose that this kind of skill-set would be required by a specialised data journalist, someone who could complete a data-based story from start to finish: from the initial idea to writing and visualising. Since teamwork skills are not present in this cluster, this journalist could presumably work alone (e.g., in a small local paper which does not have the resources to equip a big data team) or a freelancer.

Type 2: the data analyst. Cluster 2 (marked with dark blue) implies the profile of a team player, someone who works together with others to complete a project. They are skilled to locate, assess, organize and interpret various types of data using technological tools. Using these tools is closely linked to self-learning. Yet visual thinking or focusing the story are not necessarily linked to their profile. We propose this skill-set would describe a data analyst teaming up with journalists to work on a data-based story. The profile created here is very similar to the profile of German data journalists as described by Haim (2022), who also names self-learning as a notable feature. Yet what is missing compared to Haim is the visualisation side: web or graphics design. Stencel & Perry (2016) also describe a similar type: the newsroom-friendly coder.

Type 3: the "techie" of a newsroom. Cluster 3 (marked with orange) represents a journalist with advanced technical and data analysis skills. These are also linked to storytelling and a nose for news, knowing how to turn data into journalistic narratives for their audiences. Some crucial steps for gathering the data for a journalistic story are missing in this cluster: focusing the story, preparations for data gathering, choosing techniques etc. We propose that this could point to a division of tasks: while working in a data team not everyone needs to have the exact same skill-set, some jobs are done by one person, some by another. The strength of the "techie" would be visualising; not only does this journalist have visual thinking, but knows also the tools to create the graphs, charts etc. In the newsroom this person would be seen as a "techie" (Borges-Rey, 2020).

Type 4: the daily editor. Cluster 4 (marked with yellow) represents a journalist feeling comfortable working with sociological data and numerical data related to concrete areas. In this cluster crowdsourcing and interviewing stand out as the means to obtain data. Also requesting information is a skill mastered by this type of journalist, but various levels of data analysis skills are not present in this cluster. The same applies to advanced technical skills. Such journalist is not necessarily a storyteller, but rather a distributor of pre-interpreted data. In modern newsrooms such a skill-set would be required from a daily (web) editor, who needs to make quick alterations to (their own and colleague's) stories based on the growing or declining importance of a story, the developments of the events covered in a story, the spotted possibility of further audience engagement connected to the story and so on.

Type 5: the special beat editor. Cluster 5 (marked with light blue) differs from cluster 4 in the sense that this type of journalist would also be equipped with the knowledge of different data collecting methods, basic math and statistical skills, also computational thinking. This can be linked to their use of more complex data – unstructured data –, but registry data and statistics are also present in this cluster. Compared to cluster 4 a significant distinction is the skill to combine data, making it possible to come up with their own interpretations of the data. As storytelling skills are not present in this cluster, this can be interpreted as the skill-set of someone dealing with the stories of colleagues. We propose this type of skill-set would be valuable to a special beat editor, the beat being data heavy like health or economy.

Since previous studies have noted that journalists pay little attention to verifying information provided as data (Cushion et al., 2016), this skill was also coded if present in the articles (coded as *10.a Assessing data reliability*). Cluster analysis shows that the need to assess data reliability or quality was mostly discussed in terms of huge datasets. However, the need to assess the quality or reliability of data relates to other types of data too. Critical data studies show that even data collected for administrative purposes in different registers or by researchers to analyse specific social phenomena are not without bias (Gitelman, 2013; Porter, 1995). The relatively moderate frequency of this code indicates high "trust in numbers".

DISCUSSION AND CONCLUSIONS

The aim of this study was to approach data journalism in a sparsely studied way: ***a skill-based differentiation*** of data journalists. Existing academic literature was analysed and synthesised with two questions in mind: first, what kind of skills related to journalistic information processing are revealed as necessary for a data journalist and how do these skills relate to different types of data (RQ1)? And – are there associations between how data is defined by the academic literature and what journalistic skills are discussed (RQ2)? The findings aim to illustrate how academic literature (namely social sciences) has been discussing data journalistic skills but also offer a way to connect this discussion to the practical field of journalism – offering a possible match between certain data skill-sets and type of journalists using them. As a theoretical foundation, we have been building on the data concept in social sciences and data sciences (Kitchin, 2014; Raftery, 2001) in addition to the cognitive skill theory.

The corpus of literature for this study was formed searching online research databases for peer-reviewed articles: 1) on the topic of data journalism, and 2) in which the skills of producing data journalistic work were marked as key words. In total 51 relevant articles from 2008–2023 were analysed.

Five types of data-skill relations emerged from the proximity matrix created. Arguably there are factors to consider while interpreting the clusters. As limitations for this approach it has been underlined that 1) qualitative data analysis needs to be theory-driven (Keller & Achatz, 2019; Mitchell & Schmitz, 2023); 2) clustering results provide guidance for interpretation (Keller & Achatz, 2019, p. 432). That being said, based on the clusters made visible by the proximity matrix of this study, we feel confident to propose five types of media workers, each described by a slightly different skill-set, when it comes to working with various types of data. One possible way to typify them would be: ***the specialised data journalist, the data analyst, the "techie" of a newsroom, the daily editor and the special beat editor*** (RQ2).

Looking at these types through the prism of foundational vs specialist skills (Wolff et al., 2016), the balance for each type is different. For example, the daily editor has a variety of data-related foundational journalism skills, yet not so many named data specialist skills. In contrast the "techie" would also need a variety of data specialist skills. And the data analyst could lack journalistic specialist skills such as focusing a story.

Differentiating between data journalism skill-sets – and, furthermore, their foundational vs specialist skills – can prove helpful: from the viewpoint of both academia and media houses. A journalist with data-skills is sought after by employers (Stencel & Perry, 2016; Mattsson, 2020 etc), but according to Fischer's cognitive skill theory "a specific skill at one level built directly from specific skills at the preceding level" – in other words, skills must be hierarchically developed. This stresses the need for universities to rethink their "traditional trade school focus on interviewing and storytelling skills" (Nisbet & Fahy, 2015; p. 232), as journalism students would benefit from obtaining at least data confidence. At the same time, the findings of the study at hand directly address broader developments in the field of journalism: specifically, the potential for restructuring the newsrooms based on if and which data journalists are skilled to work with. While media houses worldwide are downsizing, a better understanding of which data-related skills would actually be used by journalists taking on the job of a *daily editor* compared to a *special beat editor* or *specialized data journalist*, would arguably lead to more precise job requirements on their part.

In addition, the proximity matrix presented by this study allows to conclude, that academic literature tends to define data journalism through the usage of Big Data, and in relation to skills this kind of work is not as much tied to being good at maths or statistics as to developing advanced technical skills that make it possible to perform data mining and scraping (RQ1). Yet it is evident that data journalists also need such traditional skills as having a "nose for news" and "narrating" in order to find meaningful information from Big Data and present it to the public. In conclusion, such a journalist would need to master several specialised skill-sets: those of a story-telling journalist plus those of a programmer/hacker.

Our analysis shows that the most prominent cluster of articles dealt with a broad palette of data journalism skills. Evidently, multi-skilling seems to still be expected of a journalist, making this analysis a further piece in a long line of research (e.g., Beckett & Deuze, 2016; Himma-Kadakas & Palmiste, 2019; Larrondo et al., 2016; Örnebring, 2010; Örnebring & Mellado, 2018). One way to look at innovation in journalism is through the prism of collaboration: journalists "working with data scientists, engineers, product managers to build content that is hard to replicate by other media brands" (Chew & Tandoc, 2022).

It can be summarized, that a data journalist, as most often pictured in the academic literature, is a data analyst, able to work with data in any form, a skilled narrator and a visual thinker. Although a person can acquire several specialist skills (*the "techie"*), as several of these skill-sets can indeed exist in a vertical structure, is it realistic to expect that each journalist working with data acquires multiple specialist skill-sets? This is a topic for debate: both for media houses and academia.

Liis Auväärt is a PhD student in the Institute of Social Studies at the University of Tartu, Estonia. She is also a journalist with more than 20 years of experience – currently she is a senior editor working for Estonian print media. Her research interests are datafication in journalism, journalistic skills and the relationship between data holders and the media.
ORCID 0000-0002-9786-1696
e-mail: Liis.Auvaart@ut.ee

Ragne Kõuts-Klemm is an Associate Professor in Journalism Sociology and the director of the Journalism and Communication MA programme at the University of Tartu, Estonia. In her research, she has focused on the integration of the Russian minority, changing patterns in media use and social change, journalism transparency and culture media. She has been involved in many national and international research projects, e.g., PI in research on Transparency in Estonian Journalism, 2008-2012; main investigator in the project on Culture Journalism and their Audiences, 2014-2016; researcher in a Horizon 2020 project CATCH-EyoU, Constructing Active Citizenship with European Youth: Policies, Practices, Challenges and Solutions in 2015-2018; main investigator in the project on Cross-media news consumption in European democracies, 2014-2018; PI for an applied Media Policy research project. She is actively involved in a broad international cooperation network, including the network of European cultural journals Eurozine.
ORCID 0000-0001-9277-6700

The author(s) received no financial support for the research, authorship and/or publication of this article.

No potential conflict of interest was reported by the author(s).

REFERENCES

Appelgren, E. (2017). An Illusion of Interactivity: The paternalistic side of data journalism. *Journalism Practise*, 12. DOI: 10.1080/17512786.2017.1299032

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. DOI: 10.1080/21670811.2014.884344

Asamoah, D., Doran, D., & Schiller, S. (2015). Teaching the foundations of data science: An interdisciplinary approach. DOI: 10.48550/arXiv.1512.04456

Attfield, S., Fegan, S., & Blandford, A. (2009). Idea generation and material consolidation: Tool use and intermediate artefacts in journalistic writing. *Cognition, Technology & Work*, 11: 227–239. DOI: 10.1007/s10111-008-0111-6

Ausserhofer, J., Gutounig, R., Matiasek, S., Goldgruber, E., & Oppermann, M. (2020). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, 21(7): 950–973. DOI: 10.1177/1464884917700667

Auväärt, L. (2023). Fighting COVID-19 with data: An analysis of data journalism projects submitted to Sigma Awards 2021. *Central European Journal of Communication*. 15. 379-395. 10.51480/1899-5101.15.3(32).3.

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 2053951715594634. DOI: 10.1177/2053951715594634

Battisti, M., Fedorets, A. & Kinne, L. (2023). Cognitive Skills among Adults: An Impeding Factor for Gender Convergence? CESifo Working Paper No. 10428.

Bazeley, P., & Jackson, K. (2013). Qualitative data analysis with NVivo (2nd ed.). Sage

Beckett, C., & Deuze, M. (2016). On the Role of Emotion in the Future of Journalism. *Social Media + Society*, 2(3), 205630511666239. DOI: 10.1177/2056305116662395

Beiler, M., Irmer, F., & Breda, A. (2020). Data Journalism at German Newspapers and Public Broadcasters: A Quantitative Survey of Structures, Contents and Perceptions. *Journalism Stud-*

ies, 21: 1–19. DOI:10.1080/1461670X.2020.1772855

Bhaskaran, H., Kashyap, G., & Mishra, H. (2024). Teaching data journalism: A systematic review. *Journalism Practice*, *18*(3), 722-743.

Bisiani, S., Abellan, A., Arias, Robles F., & García-Avilés, J. A. (2023). The Data Journalism Workforce: Demographics, Skills, Work Practices, and Challenges in the Aftermath of the COVID-19 Pandemic. *Journalism Practice*. DOI: 10.1080/17512786.2023.2191866

Bolaños, F., Salatino, A., Osborne, F., & Motta, E. (2024). Artificial intelligence for literature reviews: Opportunities and challenges. *Artificial Intelligence Review*, *57*(10), 259. DOI: 10.1007/s10462-024-10902-3

Bolin, G., & Andersson Schwarz, J. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data and Society*, 2(2). DOI: 10.1177/2053951715608406

Borges-Rey, E. (2020). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*, 21(7): 915–932. DOI: 10.1177/1464884917693864

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15: 1–18. DOI: 10.1080/1369118X.2012.678878

Boyles, J. L., & Meyer, E. (2017). Newsrooms Accommodate Data-based News Work. *Newspaper Research Journal* 38 (4): 428–438. DOI:10.1177/0739532917739870

Bradshaw, P. (2011). The Inverted Pyramid Of Data Journalism. *http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/* (accessed 2 April 2024).

Brechman, J. M.; Lee, C., & Cappella, J. N. (2009). Lost in Translation?: A Comparison of Cancer-Genetics Reporting in the Press Release and Its Subsequent Coverage in the Press. *Science Communication*, 30(4): 453–474. DOI: 10.1177/1075547009332649

Chaparro-Domínguez, M.-Á. & Díaz-Campo, J. (2021). Data Journalism and Ethics: Best Practices in the Winning Projects (DJA, OJA and Sigma Awards). *Journalism Practice*. Scopus. DOI: 10.1080/17512786.2021.1981773

Chew, M., & Tandoc, E. C. (2022). Media Startups Are Behaving More like Tech Startups—Iterative, Multi-Skilled and Journalists That “Hustle.” *Digital Journalism*, 12(2), 191–211. DOI: 10.1080/21670811.2022.2040374

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. DOI: 10.1080/21670811.2014.976400

Coners, A.; Matthies, B.; Vollenberg, C., & Koch, J. (2025). Data skills for everyone!(?)–An approach to assessing the integration of data literacy and data science competencies in higher education. *Journal of Statistics and Data Science Education*, *33*(1), 90-115. DOI: 10.1080/26939169.2024.2334408

Cushion, S.; Lewis, J.; Sambrook, R., & Callaghan, R. (2016). Impartiality Review: BBC Reporting of Statistics Report on qualitative research with the BBC audience. *http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/our_work/stats_impartiality/content_analysis.pdf* (accessed 2 April 2024).

Davies, K., & Cullen, T. (2016). Data Journalism Classes in Australian Universities: Educators Describe Progress to Date. *Asia Pacific Media Educator*, 26(2): 132–147. DOI:10.1177/1326365X16668969

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. DOI: 10.1080/21670811.2014.976415

Diakopoulos, N. (2019). Artificial intelligence-enhanced journalism offers a glimpse of the future of the knowledge economy. *The Conversation.* https://theconversation.com/artificial-intelligence-enhanced-journalism-offers-a-glimpse-of-the-future-of-the-knowledge-economy-117728 (accessed 2 April 2024).

Doppelt, R. (2019). Skill flows: A theory of human capital and unemployment. *Review of Economic Dynamics*, *31*, 84-122. DOI: 10.1016/j.red.2018.12.004

Engebretsen, M.; Kennedy, H., & Weber, W. (2018). Data visualisation in Scandinavian newsrooms: Emerging trends in journalistic visualisation practices. *Nordicom Review*. DOI:10.2478/

nor-2018-0007
Erkmen, Ö. (2024). Data Journalism: A Systematic Literature Review. *Journalism Studies*, *25*(1), 58–79. DOI: 10.1080/1461670X.2023.2289885
Fan, Q., Li, Y., Zhang, D., & Tan, K.-L. (2017). Discovering Newsworthy Themes from Sequenced Data: A Step Towards Computational Journalism. *IEEE Transactions on Knowledge and Data Engineering*, 29(7): 1398–1411. DOI: 10.1109/TKDE.2017.2685587
Ferrucci, P. (2018). "We've Lost the Basics": Perceptions of Journalism Education From Veterans in the Field. *Journalism & Mass Communication Educator*, 73(4): 410–420. DOI:10.1177/1077695817731870
Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States: Beyond the "usual suspects". *Journalism Studies*, 16(4): 467–481. DOI:10.1080/1461670X.2014.939852
Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological review, 87(6), 477. DOI: 10.1037/0033-295X.87.6.477
Fotopoulou, A. (2020). Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information Communication and Society*. DOI: 10.1080/1369118X.2020.171604
Garz, M. (2014). Good news and bad news: Evidence of media bias in unemployment reports. *Public Choice*, 161: 499–515. DOI: 10.1007/s11127-014-0182-2
Gehanno, J.-F., Rollin, L., Le Jean, T., Louvel, A., Darmoni, S., & Shaw, W. (2009). Precision and recall of search strategies for identifying studies on return-to-work in Medline. *Journal of occupational rehabilitation*, 19, 223–230. DOI 10.1007/s10926-009-9177-0
Gitelman, L. (Ed.) (2013). *'Raw Data' Is an Oxymoron.* MIT Press. ISBN: 9780262518284
Gond, J.-P., Mena, S., & Mosonyi, S. (2023). The Performativity of Literature Reviewing: Constituting the Corporate Social Responsibility Literature Through Re-presentation and Intervention. *Organizational Research Methods*, 26(2), 195–228. DOI: 10.1177/1094428120935494
Gray, J., Bounegru, L., & Chambers, L. (Eds) European Journalism Centre and Open Knowledge Foundation. (2012). *The data journalism handbook* (1st ed). O'Reilly Media.
Green, F. (2011). *What is Skill?: An Inter-Disciplinary Synthesis*. London: Centre for Learning and Life Chances in Knowledge Economies and Societies.
Green, S. (2018). When the Numbers Don't Add Up: Accommodating Data Journalism in a Compact Journalism Programme. *Asia Pacific Media Educator*, 28(1): 78–90. DOI: 10.1177/1326365X18766767
Haim, M. (2022). The German Data Journalist in 2021. *Journalism Practice*, 1-20. DOI: 10.1080/17512786.2022.2098523
Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism*, *18*(4), 408-424. DOI: 10.1177/1464884915620205
Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, *73*, 103– 130. DOI: 10.1016/j.euroecorev.2014.10.006
Himma-Kadakas, M., & Mõttus, M. (2021). Ready to Hire a Freelance Journalist: The Change in Estonian Newsrooms' Willingness to Outsource Journalistic Content Production. *Central European Journal of Communication*, 14(1(28)): 27–43. DOI: 10.51480/1899-5101.14.1(28).2
Hermida, A., & Young, M. L. (2019). *Data journalism and the regeneration of news*. Routledge.
Himma-Kadakas, M., & Palmiste, G. (2019). Expectations and the actual performance of skills in online journalism. *Journal of Baltic Studies*, 50(2): 251–267. DOI: 10.1080/01629778.2018.1479718
Hopkins, M. S., LaValle, S., Balboni, F., Shockley, R., & Kruschwitz, N. (2010). 10 Insights: What Survey Reveals About Competing on Information & 10 Data Points: Information and Analytics at Work. *MIT Sloan Management Review*, 52(1): 22–31.
Kallinokos, J. (2009). "On Shadows and Ghosts: Constructing Life out of DigitalRecords," Proceedings > Proceedings of ALPISitAIS, Italy. Sprouts: Working Papers on Information Systems, 9(6).http://sprouts.aisnet.org/9-6
Kashyap, G., & Bhaskaran, H. (2020). Teaching Data Journalism: Insights for Indian Journalism Education. *Asia Pacific Media Educator*, 30(1): 44–58. DOI: 10.1177/1326365X20923200
Keller, A., & Achatz, H. (2019). *Reading Between the Lines of Qualitative Data – How to Detect Hid-*

den Structure Based on Codes.
Kilpatrick, J. (2001). Understanding mathematical literacy: The contribution of research. ***Educational studies in mathematics,*** 47(1), 101–116. DOI: 10.1023/A:1017973827514
Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. ***Dialogues in human geography***, 3(3):262–267. DOI: 10.1177/2043820613513388
Kitchin, R. (2014). ***The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences.*** SAGE Research Methods. SAGE Publications Ltd. DOI: 10.4135/9781473909472
Knight, M. (2015). Data journalism in the UK: A preliminary analysis of form and content. ***Journal of Media Practice***, 16(1): 55–72. DOI: 10.1080/14682753.2015.1015801
Kosterich, A., & Weber, M. S. (2019) Transformation of a Modern Newsroom Workforce. ***Journalism Practice***, 13(4): 431–457. DOI: 10.1080/17512786.2018.1497454
Kõuts-Klemm, R. (2019). Data literacy among journalists: A skills-assessment based approach. ***Central European Journal of Communication***, ***12***(3), 299–315. DOI: 10.19195/1899-5101.12.3(24).2
Kraus, S., Breier, M., Lim, W. M. et al. (2022). Literature reviews as independent studies: guidelines for academic practice. ***Rev Manag Sci 16***, 2577–2595 DOI: 10.1007/s11846-022-00588-8
Larrondo, A., Domingo, D., Erdal, I. J., Masip, P., & Van den Bulck, H. (2016). Opportunities and limitations of newsroom convergence: A comparative study on European public service broadcasting organisations. ***Journalism Studies***, 17(3): 277–300. DOI: 10.1080/1461670X.2014.977611
Lechner, C. M., Gauly, B., Miyamoto, A., & Wicht, A. (2021). Stability and change in adults' literacy and numeracy skills: Evidence from two large-scale panel studies. ***Personality and Individual Differences***, ***180***, 110990. DOI: 10.1016/j.paid.2021.110990
Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y.; Hasshim, M. N. A. M. & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. ***British Journal of Educational Psychology***, ***82***(1), 82–99. DOI:10.1111/j.2044-8279.2010.02016.x
Lewis, N. P. & Nashmi, E. A. (2019). Data Journalism in the Arab Region: Role Conflict Exposed. ***Digital Journalism***, 7(9): 1200–1214. DOI: 10.1080/21670811.2019.1617041
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. DOI: 10.1080/21670811.2014.895504
Livingston, C., & Voakes, P. S. (2005). ***Working With Numbers and Statistics: A Handbook for Journalists***. Routledge & CRC Press. DOI: 10.4324/9781410612960
Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r) evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. ***Journalism***, 21(9): 1246–1263. DOI: 10.1177/1464884917735691
Lugo-Ocando, J., & Lawson, B. (2017). In: Nguyen, A. (ed). ***News, Numbers and Public Opinion in a Data-Driven World.*** Bloomsbury, London, pp. 62–77. ISBN: 978-1-5013-3035-3
Mattsson, A. (2020). Datafication as a Trend for Journalism: A Journalist's Perspective. In: ***Routledge Handbooks in Religion.*** Routledge. DOI: 10.4324/9780203731420-34
MaxQda (2022). ***Online Manual: Hierarchical Cluster Analysis. https://www.maxqda.com/help-mx22-stats/hierarchical-cluster-analysis*** (accessed 2 April 2024).
Miroshnichenko, A. (2018). AI to bypass creativity. Will robots replace journalists? (The answer is 'yes'). ***Information (Switzerland)***, 9(7). DOI: 10.3390/info9070183
Mitchell, G. E., & Schmitz, H. P. (2023). Using model-based clustering to improve qualitative inquiry: Computer-aided qualitative data analysis, latent class analysis, and interpretive transparency. ***VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations***, ***34***(1), 162-169. DOI: 10.1007/s11266-021-00409-8
Morini, F. (2023). Data journalism as "terra incognita": newcomers' tensions in shifting towards data journalism epistemology. ***Journalism Practice***, 1-17. DOI: 10.1080/17512786.2023.2185656
Nisbet, M. C., & Fahy, D. (2015). The Need for Knowledge-Based Journalism in Politicized Science Debates. ***The ANNALS of the American Academy of Political and Social Science, 658(1),*** 223–234. DOI: 10.1177/0002716214559887
Ojo, A., & Heravi, B. (2018). Patterns in Award Winning Data Storytelling. ***Digital Journalism***, ***6***(6), 693–718. DOI: 10.1080/21670811.2017.1403291

Örnebring, H. (2010). Technology and journalism-as-labour: Historical perspectives. *Journalism*, 11(1): 57–74. DOI: 10.1177/1464884909350644

Örnebring, H., & Mellado, C. (2018). Valued skills among journalists: An exploratory comparison of six European nations. *Journalism*, 19(4): 445–463. DOI: 10.1177/1464884916657514

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. DOI: 10.1177/1461444812463345

Parti, K., & Szigeti, A. (2021). The future of interdisciplinary research in the digital era: Obstacles and perspectives of collaboration in social and data sciences-an empirical study. *Cogent Social Sciences*, *7*(1), 1970880. DOI: 10.1080/23311886.2021.1970880

Persaud, A. (2020). Key competencies for big data analytics professions: A multimethod study. *Information Technology & People*, 34(1): 178–203. DOI: 10.1108/ITP-06-2019-0290

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide* (pp. xv, 336). Blackwell Publishing. DOI: 10.1002/9780470754887

Porter, T. M. (1996). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.

Portilla, I. (2018). Contributions of academic articles to the practice of journalism and data management. *Observatorio (OBS*)*, 12(3). DOI: 10.15847/obsOBS12320181245

Raffaghelli, J. E., & Stewart, B. (2020). Centering complexity in 'educators' data literacy' to support future practices in faculty development: a systematic review of the literature, *Teaching in Higher Education*, 25(4): 435–455. DOI: 10.1080/13562517.2019.169630

Raftery, A. E. (2001). Statistics in sociology, 1950–2000: a selective review. *Sociological Methodology*, *31*(1), 1-45.

Ramsälv, A., Ekström, M., & Westlund, O. (2023). The epistemologies of data journalism. *New Media & Society.* DOI: 10.1177/14614448221150439

Reich, Z., & Godler, Y. (2014). A Time of Uncertainty. *Journalism Studies*, *15*(5): 607–618. DOI: 10.1080/1461670X.2014.882484

Rodríguez, J., & Clark, A. (2021). Big Data and Journalism: How American Journalism is Adopting the Use of Big Data: How American Journalism is Adopting the Use of Big Data. *Novum Jus*, 15: 69–89. DOI: 10.14718/NovumJus.2021.15.1.4

Rowe, F. (2023). *Big data.* Concise Encyclopedia of Human Geography (pp. 42-47). Edward Elgar Publishing. DOI: 10.4337/9781800883499.ch09

Royal, C. (2012). The Journalist as Programmer: A Case Study of The New York Times Interactive News Technology Department. *ISOJ Journal, Volume 2, Number 1, Spring*, 5–24.

Sirén-Heikel, S., Kjellman, M., & Lindén, C.-G. (2023). At the crossroads of logics: Automating newswork with artificial intelligence—(Re)defining journalistic logics from the perspective of technologists. *Journal of the Association for Information Science and Technology,* 74(3): 354–366. DOI: 10.1002/asi.24656

Sharma, S. (2017). Definitions and models of statistical literacy: a literature review, *Open Review of Educational Research*, 4:1, 118-133, DOI: 10.1080/23265507.2017.1354313

Showkat, D., & Baumer, E. P. (2021). Where do stories come from? examining the exploration process in investigative data journalism. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2): 1–31. DOI: 10.1145/3479534

Spry, C., & Mierzwinski-Urban, M. (2018). The impact of the peer review of literature search strategies in support of rapid review reports. *Research Synthesis Methods*, 9(4): 521–526. DOI: 10.1002/jrsm.1330

Stalph, F. (2017). Classifying Data Journalism: A content analysis of daily data-driven stories. *Journalism Practice*, 12: 1–19. DOI: 10.1080/17512786.2017.1386583

Stalph, F., & Borges-Rey, E. (2018). Data Journalism Sustainability. *Digital Journalism*, 6(8): 1078–1089. DOI: 10.1080/21670811.2018.1503060

Steensberg, M. T. (2021). In Numbers We Trust – The Use of Statistics in Political News. *Journalism Practice*, 0(0): 1–19. DOI: 10.1080/17512786.2021.1987298

Stencel, M., & Perry, K. (2016). Superpowers: The digital skills media leaders say newsrooms need going forward (United States of America) [Report]. *Tow-Knight Center for Entrepreneur-*

ial Journalism. https://webdesign.cindyroyal.net/handouts/Superpowers_%20Digital%20Skills%20Newsrooms%20Need_.pdf (accessed 1 April 2024).

Thurman, N., Dörr, K., & Kunert, J. (2017). When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences. ***Digital Journalism***, 5. DOI: 10.1080/21670811.2017.1289819

Treadwell, G., Ross, T., Lee, A., & Lowenstein, J. K. (2016). A Numbers Game: Two Case Studies in Teaching Data Journalism. ***Journalism & Mass Communication Educator***, 71(3): 297–308. DOI: 10.1177/1077695816665215

Vahey, P., Rafanan, K., Patton, C., Swan, K., van't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. ***Educational Studies in Mathematics***, ***81***, 179–205. DOI: 10.1007/s10649-012-9392-z

Van Der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. ***Proceedings of the National Academy of Sciences***, ***117***(14), 7672-7683. DOI: 10.1073/pnas.1913678117

van Dijck, J. (2016). Foreword. In: Schäfer, Mirko T., & van Es, K. (eds), ***The Datafied Society. Studying Culture through Data***. Amsterdam University Press. DOI: 10.5117/9789048531011

Van Witsen, A. (2020). How Daily Journalists Use Numbers and Statistics: The Case of Global Average Temperature. ***Journalism Practice***, 14(9): 1047–1065. DOI: 10.1080/17512786.2019.1682944

Veglis, A., & Bratsas, C. (2017). Reporters in the age of data journalism. ***Journal of Applied Journalism & Media Studies***, 6: 225–244. DOI: 10.1386/ajms.6.2.225_1

Wagner, G., Lukyanenko, R., & Paré, G. (2022). Artificial intelligence and the conduct of literature reviews. ***Journal of Information Technology***, ***37***(2), 209-226. DOI: 10.1177/02683962211048201

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. ***The Journal of Community Informatics***, ***12***(3). DOI: 10.15353/joci.v12i3.3275

Wu, S. (2024). Assessing the growth of data journalism start-ups as alternative media and their roles in "flawed democracies". ***Journalism***, 25(6), 1310–1327. DOI: 10.1177/14648849231177830

Wu, S., Tandoc, E. C., & Salmon, C. T. (2019). A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency. ***Digital Journalism***, 7(4): 428–446. DOI: 10.1080/21670811.2019.1620112

Young, M. L., Hermida, A., & Fulda, J. (2018). What Makes for Great Data Journalism? ***Journalism Practice***, ***12***(1), 115–135. DOI: 10.1080/17512786.2016.1270171

Zamith, R. (2019). Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. ***Digital Journalism***, 7(4): 470–489. DOI: 10.1080/21670811.2018.1554409