*Article*

# The epistemologies of data journalism

## Amanda Ramsälv 🆔
## Mats Ekström
University of Gothenburg, Sweden

## Oscar Westlund
Oslo Metropolitan University, Norway; University of Gothenburg, Sweden

## Abstract
Amid digital developments, data journalism has gained a strong foothold among news publishers and in public discourse. With its authoritative claims and informative visualizations, it can play a significant role in the actions of citizens and people in power. This mixed-method case study explores a distinct epistemology developed in an independent form of data journalism in public service media in Scandinavia, not subordinate to traditional news values or investigative journalism. The study investigates its knowledge and truth claims, approach to data, transparency practices, and resources invested to claim reliable knowledge. The epistemology is characterized by innovative practices in the visualizing of essentially prejustified datasets. It claims public value offering general information and audience-friendly explorations of individual perspectives on topics on the public agenda. The approach to data views reality as measurable facts yet indicates epistemic ambiguity regarding figures' reliability, guided by a principle of reasonableness in the justifications of truth claims.

## Keywords
COVID-19, data journalism, epistemology, forms of knowledge, transparency, truth-claims

**Corresponding author:**
Amanda Ramsälv, The Department of Journalism, Media and Communication, University of Gothenburg, Seminariegatan 1 B, Gothenburg 41313, Sweden.
Email: amanda.ramsalv@jmg.gu.se

## Introduction

The developments within news journalism have significant implications for epistemologies; what is claimed to be known, the acquiring and processing of information, and the practices and standards applied in justifications of truth claims. As knowledge production is socially conditioned and variable within and beyond traditional genres of news journalism, the epistemologies are marked by diversity (Ekström, 2002; Ekström et al., 2021; Kligler-Vilenchik and Tenenboim, 2020; Matheson and Wahl-Jorgensen, 2020). Innovations within news genres also claim to contribute different forms of knowledge to the public (Park, 1940). This is true not least for data journalism. This study investigates the epistemology of a particular form of data journalism distinct from the use of data and visualizations in daily news and investigative reporting (Borges-Rey, 2016), which thus ultimately stretches the approach to knowledge within journalism.

Data journalism enables journalism to develop the production of public knowledge in essentially different ways than possible in other forms of news reporting. In the processing and visualizations of data with consistently strong truth claims about public information, data journalism adds approaches to knowledge in journalism (Appelgren et al., 2019; Borges-Rey, 2020; Splendore, 2016). A key reason is that data journalism is shaped by the merging of norms, standards, and practices originating from epistemologies in different institutional contexts (news journalism, computational and academic). Data journalism thus introduces various epistemological suggestions into journalistic practices (Parasie and Dagiral, 2013), "mediated by a constant interplay between an approach that is deeply rooted in journalistic conventions and an approach that is increasingly reliant on computational processes and logics" (Borges-Rey, 2020: 920). This applies, for example, to how journalists gather and analyze data to draw conclusions (Coddington, 2015: 341) and whether data are considered to have value in themselves or only in relation to news stories (Borges-Rey, 2020; Parasie, 2015). Ultimately, the different forms of data journalism have distinct yet evolving characters, and their implications for epistemology are thus important to study (Royal and Blasingame, 2015).

In this article we investigate a form of data journalism characterized by four distinctive features: (1) the relative independency from traditional genre conventions of news or investigative journalism (Borges-Rey, 2016), and the related news values (Tandoc and Oh, 2017), and claims of scrutinizing those in power, (2) the predominantly usage of publicly available datasets produced by pre-justified sources, (3) the emphasis on providing opportunities for audiences to adjust visualizations to be relevant for their specific situations, and (4) the striving for making data processing transparent and intelligible for audiences through meta information such as method sections. These four features per se are not unique for this form, as several of them form core features for data journalistic practices in general. Yet the composition of these four together characterizes a form of data journalism distinctive from data journalism focused on in previous research (Borges-Rey, 2016). We have studied a team engaged in this form of data journalism, in Swedish public service media (PSM). We hereafter refer to this as *independent data journalism* (IDJ).

A definition of the term data journalism is in itself under scholarly discussion (Appelgren and Nygren, 2014; Coddington, 2015). We define data journalism following

Coddington (2015: 334): "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." More specifically, we use *forms of data journalism* when data journalistic tools are used in different journalistic editorial contexts, in line with the division of Borges-Rey (2016: 838): "daily, quick turnaround, generally visualised, brief forms of data journalism; extensive, thoroughly researched, investigative forms of data journalism; and light, editorialised, entertaining, often humorous, gamified forms of data journalism." The literature to date shows research on epistemologies in different forms of data journalism. First, when data are used in the processing of large datasets in advanced forms of investigative journalism (Young et al., 2018), where Parasie (2015: 373) recently studied the epistemological tensions when journalistic conventions of truth claims are paired with using data in the process of justifications. The study shows how data processing is applied to assist this form of journalism both following a "hypothesis-driven approach" where data are checked in relation to certain assumptions or a "data-driven approach" where data are embraced without specific hypothesis but instead "expect from data processing the identification of new and unexpected stories" (Parasie, 2015: 376). Borges-Rey (2020) distinguished researchers' approach to data journalism into either epistemological emphasis following a "newshound" or "techie" approach to study data journalism in the United Kingdom, concluding that "data journalists displayed a blend of journalistic and computational skills as they moved back and forth between the newshound and the techie approaches" (Borges-Rey, 2020: 929).

This study contributes to the literature on epistemology and data journalism in two main ways. First, it presents a sociological approach to the study of epistemologies. While epistemology is brought into detailed analyses of different forms of journalism, only a few studies have systematically addressed epistemology within data journalism (e.g. Borges-Rey, 2020; Parasie, 2015; Stalph, 2018). Second, it advances knowledge about IDJ as a distinct form of data journalism highly independent from both genre conventions of news reporting and investigative journalism, a form that indeed demonstrated its critical impact on public discourse and knowledge during the COVID-19 pandemic (Desai et al., 2021; Pentzold et al., 2021; Wu, 2021). Through analyzing the everyday data journalistic work operating through the four abovementioned features from the perspective of news epistemology, we claim a distinct contribution to the literature on news and its diverse epistemologies (Kligler-Vilenchik and Tenenboim, 2020; Matheson and Wahl-Jorgensen, 2020). The article offers a mixed-methods case study of the team engaged in IDJ, investigating the epistemology through the following four research questions: RQ1: *What characterizes the form of knowledge?*, RQ2: *How are data understood and approached?*, RQ3: *How are practices of transparency applied?*, and RQ4: *How are epistemic efforts calculated and balanced?*

## Data journalism and epistemology

Data journalism involves diverse technologies and datasets (Ojo and Heravi, 2018), and data analysis occurs as isolated visualizations or alongside text and audiovisuals (Coddington, 2015; Loosen et al., 2020; Splendore, 2016; Zamith, 2019). Data journalism emerged in the 20th century with computer-assisted reporting and precision

journalism (Anderson, 2018). Various researchers have studied emerging forms of data journalism in the Global South (de-Lima-Santos and Mesquita, 2021; Lewis and Al Nashmi, 2019; Mutsvairo et al., 2019; Palomo et al., 2019) and the global north (Appelgren and Nygren, 2014; de Maeyer et al., 2015; Fink and Anderson, 2015), where an increased dialogue in between scholarship in different parts of the world is suggested (Wright et al., 2019). Scholars have contributed valuable research overviews of data journalism studies (Cheruiyot et al., 2019; Heravi et al., 2022; Splendore, 2016). In line with the purpose of the study, the literature review hereunder focuses on research into the intersection of data journalism and epistemology, related to our four research questions.

This article adopts a sociological perspective on epistemology, studying forms of knowledge, knowledge-producing practices, routines, and standards within social contexts (Ettema and Glasser, 1987). The diversification of news journalism creates essentially different epistemologies: What it claims to know about, how it knows, and how these claims are justified (Ekström et al., 2021; Kligler-Vilenchik and Tenenboim, 2020; Matheson and Wahl-Jorgensen, 2020).

Journalism claims to provide important public knowledge, a task taking different forms. In a seminal study, Park (1940) situated news as a form of knowledge in "a location of its own," on a continuum between knowledge as "acquaintance with" and "knowledge about," enabling ordinary people easily to understand and talk about their world (p. 675). As a form of knowledge, news arguably includes different values for people, assumed and claimed by providing news in specific ways (Ekström et al., 2022). The approach to news as *one* form of knowledge is extended in a diversified digital mediascape (Nielsen, 2017). Previous data journalism research has addressed reporting ambitions in different ways. Borges-Rey (2020) notes how data journalists are providing data in a traditional journalistic way or as an emergent approach with more computational data operationalization. In a study of Scandinavian newsrooms, Engebretsen et al. (2018: 9) noted that the claim of empirical evidence is one main function motivating journalists to prioritize data visualizations. Parasie (2015: 365) discussed this as a "renewed promise of objectivity," which can be seen as another assumed value. A fundamental issue actualized in current research is the extent to which data journalism complements traditional news journalism's approach to knowledge production or constitutes something radically different. In a study of big data journalism and its relation to traditional news values, norms, and routines, Tandoc and Oh (2017: 997) conclude that "big data journalism shows new trends in terms of how sources are used, but still generally adhere to traditional news values and formats such as objectivity and use of visuals." This article posits IDJ as a distinct form of knowledge, produced and presented without a professional gut feeling of traditional news values (Schultz, 2007) by multiprofessionals making varying epistemological suggestions (Borges-Rey, 2020; Parasie and Dagiral, 2013). *RQ1* investigates this topic suggesting IDJ offers possibilities for audiences to shift between general and individual forms of public knowledge.

News journalism is an established fact-based discourse (Ekström and Westlund, 2019). Data journalism's foundation pillar for claiming truth is data processing and presentation. Research on approaches to data, central to the epistemological account of making truth claims using different sources of information, has highlighted data journalism's lack of routines in critically assessing datasets, resulting in it essentially carrying dataset's quantitative discourses (e.g. Lugo-Ocando and Nguyen, 2017). Borges-Rey (2020)

discussed how data journalists acquiring public data treat them as mere "data" rather than others' work output, with inherent limitations, but nevertheless attempt to identify weaknesses and transform them into a format with which they can work. Lawson's (2021) study investigating how UK journalists covered humanitarian crises showed that, rather than checking and verifying data materials, journalists opted to select carefully the sources from which they acquired datasets. In *RQ2*, we focus on the understanding of data in the production of visual components, different methods of visualization as a pivotal part of data journalism, and their presentation to audiences in the text as representations of reality.

In the context of truth claims, transparency is central. Lewis and Waters (2018: 730) addressed different journalistic methods, arguing that for audiences "Data journalism is a vague and less familiar term that potentially could nourish audience fears of behind-the-scenes manipulation." This is compared to methods such as interviewing that makes sense merely through the consumption of news. Relevant for truth claims is also the broader discussion of transparency as a means of enhancing journalism's democratic potential when data journalists process digital documents with data and make them intelligible for the audiences (Parasie, 2019). In our analysis, we have identified transparency as central to the development of epistemic practices in IDJ. This brings in the broader transparency literature (Karlsson, 2010) to help analyze a significant part of epistemic practices in IDJ. *RQ3* investigates the varying levels of transparency associated with work toward making truth claims.

Knowledge-producing practices are shaped by local agreements about what is necessary or important to prioritize when striving for (internally) acceptable reasons to be able to claim knowledge, including information verification (Godler and Reich, 2017). Specific agreements about epistemic efforts shift from context to context (Ekström et al., 2021) in what Parasie (2015: 373) described as "the collective production of justified beliefs about the world" within particular news production contexts. Research has established that publishers perceive certain sources as legitimate (Cushion et al., 2017; Ettema and Glasser, 1987) and decide when data have undergone sufficient checks to be published (Lawson, 2021). In *RQ4*, the study synthesizes the analyses and provides an additional layer, focusing on how variations in investment in the different epistemic practices are decided.

## The case, methods, and material

### The case

We study a case of the form of data journalism (IDJ) discussed earlier. It concerns a data journalism team (DJT) at Swedish public service television (SVT). The team consists of five media workers with various professional backgrounds (journalism, engineering, public information, and digital development). Some media workers are employed by the news division and some by the IT. The DJT works rather independently regarding content and focuses on experimenting and developing data processing and visualizations. The DJT has been organized so that the team can decide on their own projects and prioritizations in everyday work, albeit coordinating with the news division. Also, the DJT is rather autonomous in that they publish and edit their publications for a distinct news site URL (separated from general news division). This enables them to independently

publish, prioritize, edit, and control the publications on this site. The team publishes on a wide range of topics, such as COVID-19, political elections, crime, weather, etc.

## Methods and material

The study was carried out in 2020–2021 following a mixed-methods approach:

1. Semi-structured interviews with four (out of five) DJT members (not the person responsible for IT and web infrastructure), conducted by two of the researchers. The interviews addressed the central epistemological aspects guiding the study: what is claimed to know something about, how is it known and how are these claims justified in everyday work, focusing on the routines and practices applied in the acquiring and processing of information. We also conducted two interviews with the head of strategy, setting the organizational context, as well as an interview with the chief data journalist in the PSM newsroom, focusing on producing data journalism for news journalists and occasional investigative data journalism projects.

2. A preliminary analysis by all three researchers of DJTs around 100 published reports focusing on the epistemological aspects related to in our research questions, which served to inform:

3. Repeated interviews with the four DJT members, and for two also a third interview (a total of 10 interviews, 889 minutes of recording, with DJT members), asking more specific questions (including also shared screens guiding us through the different stages).

4. The interviews and preliminary analysis developed the focus for a manual in-depth qualitative analysis of 22 of their publications (June 2021 by one of the researchers). The reports were selected to cover a variety of topics, with respect to also topics that are covered through many publications and more occasional and topics that are central and more peripheral to the news agenda. These selection criteria contributed to rich data for the analysis in enabling a variation representative of the reports published by the team. The analysis focused on the headline, lead paragraph, data updates, type of visualizations, (preset) information visualizations, source(s), method selection, salience of the method in publication, information production of the dataset, possibilities for audiences' exploration, discursive resources' truth-claims and logs of changes. The researcher carefully took notes for each of the aspects.

5. In this step, the three researchers synthesized the results from both the interview material and analysis of the publications, in relation to the four research questions.

## Findings: IDJ

### What characterizes the form of knowledge? (RQ1)

This section addresses the form of knowledge claimed by the DJT, indicating the assumed values for audiences. With an explicit ambition not to follow traditional news values and

by experimenting with data visualizations, the DJT claims a form of knowledge where audiences can navigate publications between two forms of knowledge, general and individualized. This forms news assumed to be relevant without being adapted to (and framed by) strict news values such as timeliness, negativity, and unexpectedness.

*News can be slow and unspecified.* The DJT members come from various professional backgrounds (Parasie and Dagiral, 2013). This specialist team also has a rather exclusive position in its organization, having autonomy in developing innovative visualizations. This combination arguably enables less restricted thinking about what news is and should be, in the absence of traditional news values and gut feelings, which otherwise guide news production.

DJT explicitly challenges the widespread ideal of fast-and-first content 24/7 (Usher, 2018). The topics often have some actuality factor, albeit from a wider scoop perspective than justified as adding to a particular news agenda, instead addressing—what are current topics and processes in public discourse? An example is the publication "Checking out the carbon" before the upcoming climate meeting. In addition to what to report, DJT members also express an ambition to remove the journalistic window to the world in defining specific perspectives and foci. A DJT member explained the reluctance to use filters:

> So that we don't pick out three numbers to tell you about, but here you can see for yourself and you can compare yourself if you want to. . . . there are no, not so many filters in between.

In traditional news reporting, public knowledge is provided to audiences following news values of what makes a good news story (such as timeliness, negativity, and unexpectedness). DJT departs from other criteria and claims and indicates to the audiences instead forms of knowledge that are (1) *general* and (2) *individualized*, on the same topic. Through the usage of different data journalistic tools, they give the possibility for audiences to explore and move back and forth between these two forms.

1.  General knowledge

In headlines and lead paragraphs, no specific aspect of the chosen topic is in focus, the presentations indicate that general information regarding the topic is included. In a publication on patients in intensive care and inpatient care, the headline is "Corona in intensive care and inpatient care," indicating that no information within the article is more worthy of attention. Conversely, it highlights the assumed value of providing something general, with no particular angle, and that of providing information about all patients and regions.

2.  Individualized knowledge

While news journalism essentially represents a one-for-all form of news publishing, publishers have developed other approaches to digital journalism. Data-driven news work is geared toward producing and publishing the news that audiences want using analytics

and metrics extensively (e.g. Ekström et al., 2022). As an extension, publishers have developed technology-driven personalized news systems feeding users with different news diets (e.g. Helberger, 2019). The DJT offers a form of personalization through what we refer to as individualized knowledge enabling audiences to explore visualizations from their individual angles, yet not using these forms of data-driven and personalized practices. The individual angles do not involve knowing something unique about persons but use, for example, their geographic location (for everyone to connect to locally). This is similar in character to information historically included in the news, in the reporting of sports results, forecasts, and so on. In this way old reporting techniques are simply actualized in a modern digital editorial context (Jacobson et al., 2015). What really differs however is the hyperlocal framing and encouragements, such as "Reported crimes where you live," "Certificated teachers in your school," and "The divide in income in your municipality," in relation to how it also claims general information. In a publication about vaccinations, the headline is "The corona vaccine in Sweden and the world." After reading numbers on the country level, the digital design enables audiences to choose "your region and municipality" to obtain information about the vaccination in specific places in Sweden, and a pointing hand symbol on a digital map over Sweden encourages the reader to "Click on the map" to get this local information. The map indicates percentages of vaccinations with different colors. In line with how Usher (2020) connects digital tools to distinct shaping of knowledges, in arguing that the digital map serves to "offer new ways to construct knowledge via information exploration," the newsworthy information(s) in the digital design with possibilities for audience exploration is what the audiences themselves choose to explore as knowledge (p. 252). However, arguably with a tendency of paternalistic features in explicitly encouraging the audiences to explore this news value, through the interactive design in the publications (Appelgren, 2018). This actualizes Splendore's (2016: 349) discussion about data journalism as an individual service, and service journalism in the covering of typically hard news (Widholm and Appelgren, 2022).

## How are data understood and approached? (RQ2)

Journalism that mainly represents the world in numbers, statistics, graphs, charts, and maps may create an aura of facticity and objectivity within the team as well as in relation to the audiences (Engebretsen et al., 2018). Numbers signify exactness, and their rhetorical power lies in the viewers' opportunity to see them with their own eyes. In the visualization of data, IDJ relies heavily on figures as pure facts and the actual data production tends to be disregarded. However, this is not to suggest a simple and general approach to figures. On the contrary, this section shows contrasting understandings of data within the IDJ culture and practices. These are also further contextualized in the analyses of transparency practices and prioritizations of epistemic efforts (RQ3 and RQ4).

*Figures show how it is.* The team enacts and cultivates particular approaches to data through the construction of articles. In the significant relationships between text in headlines and lead paragraphs and the forms of visualizations, we identify recurrent ways of

representing data: (1) text refers to how evidence can be *seen* in the figures and (2) data (as figures) are represented as *independent* evidence without reference to the production of data or agency in interpretations.

In the articles, the headlines and lead paragraphs typically point forward to figures in the visualization rather than summarizing a news story. The lead paragraph in an article on climate change for example states: "Here you can follow the development of global warming." (Note the forward-referring spatial deixis, "here," and the invitation to "follow"). The article represents the development, among other things, with a counter showing how we are approaching zero in the global carbon dioxide budget (in relation to the goal of 1.5 degrees). The artifact used in the visualization indicates precision and accuracy.

Cartography and the related mapping of reality are among the most frequent and institutionalized forms of visualization in data journalism and are often used by the team. Cartography is the ultimate form of the relationship between offering seeing in the headline or lead paragraphs and including figures as observable facts. A map is a conventional and familiar way of showing facts, actualizing a natural and unquestionable discourse. In the study a publication about refugees across the world, the lead paragraph is:

> Since 2000, 21 million people have fled wars and conflicts around the world to find shelter in other countries. 752,638 of them have come to Sweden to seek asylum. Scroll to see where the refugee flows come from—and where they go.

The publication includes a globe with moving lines illustrating a constant flow of refugees toward Sweden. Wordings throughout the publication, such as "this is what it looks like" and "now we look at," in relation to the accounts of numbers and maps, indicate pure facts in a natural representation of reality. However, the relationship between lexical choices and visualization (lines in constant motion directed toward Sweden from different parts of the world) at the same time reproduces a particular representation of refugees as constant flows, differently sized, toward the reader's assumed location. As Usher (2020) noted, maps are journalistic resources claiming to make the complex comprehensible for the reader. Moreover, as the example above indicates, the reduction of complexity categorizes actors from certain perspectives and presents narratives of place and socio-cultural relationships (Usher, 2020: 258).

*Ambiguity in the approach to figures as facts.* The DJT members indicate figures' fundamental ambiguity as facts. They express a strong belief that numbers show facts, as described in the examples above. Numbers are attributed epistemic qualities with advantages over words, in being less loaded with values and emotions. Numbers can be wrong but also corrected. Specifically, numbers are associated with a clear distinction between accuracy and incorrectness, ultimately between true and false. Practices are applied to ensure that publications include as few errors as possible. This understanding of figures is indicated in this interview quote in which a DJT member reflects on their work:

> It's all clearly verifiable, often. It is not so emotionally driven as it can actually be in other journalism. These are actual figures and we are terrified of being wrong. . . . It happens of course that we are wrong and then we correct and say thank you so much for informing us that we were wrong. But it's not fun.

However, DJT members also express an awareness of data's uncertainty, problems, and limitations. They have extensive experience with the challenges of using available data to present a reliable picture of reality. One says: "It is difficult to present correctly." They know (not least based on their own experience) that data are produced in several steps and by various actors, with quality implications. They are aware of the problem of often having to rely on data without knowing how they are produced. This creates ambiguity in the understanding of data as a means of representing reality. Nevertheless, the conventional form for the relationship between text and visuals in their publications communicates figures as observable evidence, independent from production and interpretation practices.

## How are practices of transparency applied? (RQ3)

Information about the production process allows audiences to understand and criticize the news (Kovach and Rosenstiel, 2001). Data journalism promises to make even the world itself more transparent (Parasie, 2019: 264), rendering the increasing amount of digital data (produced by governments and others) intelligible to audiences. This section addresses the question of how practices of transparency are structured and prioritized in the DJT's publications. The analysis shows how the DJT applies different transparency practices and how their amount and salience vary on a spectrum from minimal to extended ambitions.

*Information about the dataset and reading instructions.* Through content analysis of the DJT's publications, we identify two practices of transparency as central to what Karlsson (2010: 537) labeled "disclosure transparency": *information about the dataset* and *reading instructions.* These practices are widely used in publications. The practice of providing *information about the dataset* can involve either information about the source and data set accompanied by a link to the actual dataset or clear descriptions of the data. An example is the description of the data used for the publication about the wolf population in Sweden: "The estimated number of wolves comes from the inventory reports that are done every year by Rovdata, Viltskadecenter and Høgskolen i Innlandet." Audiences are thus informed about the data used and may search for the actual dataset. The focus is on the dataset per se and not its production. The broader discussion of data journalism as a promise to provide transparency for citizens (Parasie, 2019) is actualized here in ultimately asking if investigating and presenting the producing practices of datasets is included in the practices of making digital documents intelligible for the public.

Providing *reading instructions* is another key practice of DJT. While not necessarily adding anything that was previously missing, the instructions underline specific aspects. The following example is from a publication about how the team handles data on

COVID-19 deaths when the Public Health Agency sometimes has to adjust its numbers in hindsight:

> If we look closely at the red line while scrolling, we can see how it can raise in hindsight, even rather far back in the graph, so, what this means is that the Public Health Agency has adjusted the numbers for a day far back in time.

The graph described in the example follows the text explaining that it represents the difference between the first information and the adjusted definitive numbers. Nevertheless, the example above is from the text that appears immediately after the graph. It is not new information that adds or alters something: "so, what this means is that." However, it underlines the already existing information. It also invites collective reading—"If we look closely (. . .) we can see"—functioning as pedagogic in holding the readers' hands.

*A spectrum from minimal to extended transparency.* The publications recurrently follow the two practices of providing information about the dataset and reading instructions. However, the extent to which the DJTs engage in different transparency practices varies. We show this by focusing on two poles of a transparency spectrum illustrated through two publications: minimal and extended. *Minimal* transparency is exemplified through the few brief transparency practices in the publication "The sun league 2021," reporting on Sweden's hours of sun. Methods and datasets or sources are mentioned merely in the following text at the end of the publication: "The sun league 2021 is a cooperation between SVT Weather and SVT Data journalism." In the publication, there is transparent information on the source of errors, where the DJT use disclaimers regarding measurements and comparisons. However, the problems are not presented as crucial, although the publication's key content aims to enable an exploration and comparison of different years of sun hours across Sweden.

The COVID-19 case is particularly challenging for data journalists due to the interest in and prioritizing of, for example, comparisons of numbers between different countries, with variations in both quality and forms of collecting and producing data (Desai et al., 2021). With strong truth claims, these complicating factors are part of everyday work. In this context, we identify three overarching and distinct practices of transparency in the main COVID-19 article, "The spreading of the new corona virus," pushing this to be an example of *extended* transparency. These are: (1) *Emphasize methods, (2) make disclaimers salient*, and (3) *structure through time*. The practices themselves contain different subpractices, exemplified below.

*1. Emphasize methods*

> In "The spreading of the new corona virus," the emphasizing of methods is represented through communicating information about methods as *key information* in relation to the content. There are explicit encouragements to visit the methods section, such as immediately after the lead paragraph: "Read the texts around the charts carefully, and see the method section at the bottom of the article to get it all right." The emphasis on methods is apparent also through the

presentation of a plurality of transparency practices in a methods section at the bottom, with its own headline: "Source and method."

*2. Make disclaimers salient*

Another overarching practice of transparency is making disclaimers *salient*, contrary to dutiful mentioning placed in the background. The quality of the information is indicated by openness about knowledge limitations. Disclaimers are treated as central to the audiences' possibility to understand the content correctly. An example from the material is when a lead paragraph of a publication consists solely of the following disclaiming information:

> The numbers in this article only show the confirmed cases of infected, deceased and recovered people. Different countries have different routines for if and when they test people and are likely to have a large number of unreported cases not only for the number of infected people but also for the recovered ones.

*3. Structure through time*

Some of the publications have evolved and adjusted in a constant process during a long time period. The publication in focus was published in January 28, 2020 and still updated May 2022. This evolving state was structured epistemically by informing audiences *when* things of importance happened. The DJT states when numbers were updated and why: "The numbers in this section focusing on Sweden are updated once a day Tuesday to Friday when the Public Health Agency releases their daily report—around 14:00." In this publication exemplifying extended transparency, the DJT includes a separate section labeled "log of changes," with *timed* changes such as for example: "2020-04-06: The figure of deaths in each country is expanded with the opportunity to compare more countries." By providing a time structure, the DJT enables the audiences to understand how this constant publication (and construction of knowledge) develops when publications are less predictable than other forms of news reporting.

## How are epistemic efforts calculated and balanced? (RQ4)

A critical aspect of epistemology concerns the epistemic efforts invested to justify truth claims (Ekström et al., 2021) where a prioritized practice in the DJT's work is to do various forms of checks for errors in the data and visualizations. The study further suggests two concrete calculating factors defining the level of different epistemic efforts: (1) *What is the source?* and (2) *What is the topic?* In addition, we identify the overarching principle of *reasonableness* to help balance the efforts to make them feasible for operation in practice.

1. *Source.*

Journalists rely heavily on sources perceived as credible (Reich, 2011). The epistemic efforts undertaken by the DJT decrease when using a credible, prejustified source (Ettema and Glasser, 1987). When reporting on considered serious topics, they often use public sources, such as governmental agencies and universities. These are prejustified;

any further justification is not perceived as necessary. The checking of information is therefore not extended, ultimately creating an "epistemological debt" (Usher, 2020: 252). A DJT member explicates their reasoning about data from a governmental agency:

> But if one cannot use statistics from the National Agency for Education, then I don't know. That is used by politicians and everything so that has to be okay to show.

For these cases, the epistemic efforts are built into the legitimacy of the organizations themselves (Lawson, 2021). There is a shared understanding within the DJT regarding which sources that belong to this group. Epistemic efforts also decrease when sources are not initially perceived as prejustified but achieve this status through having been accepted by others deemed credible. In this way, the credibility of that other party can in itself transform a source into a legitimate source. A DJT member explains the implications of other media workers using a source:

> But if for example New York Times would have used it, then I feel more, it feels more like a legit source, than if I only find it on Hasse's blog, who has done his thing.

2. *Topic.*

It is evident from our interviews that some topics are considered worthy of more epistemic efforts than others. A DJT member explains this calculation in response to a question on how the DJT handles the usage of commercial sources:

> . . . It also depends, to be honest, on what topic it is. . . . it depends on you know how important it is or how, how important the topic is, how important it is that the data is correct.

It is also considered somewhat self-evident which topics should not receive as much attention, as another DJT member explains when talking about various levels of importance:

> . . . But one of the questions was only the price for beer in different countries, . . . but then it is you know like this, what is the cost of being wrong about this? . . . but I guess it is much worse to do that if you are to say this many think we should welcome less immigrants . . . you know it feels like a worse, just a worse question to be wrong about.

This way of describing topics, using wordings of self-evidence such as "you know" (Sw "ju"), means that further argumentation is not needed. However, some topics that the DJT probably considers serious are actually not subject to extended epistemic efforts. The publication "The issues that divide Sweden" builds on data from a poll on topics such as taxes, the environment and immigration. The headline makes a strong claim, yet, at the end of the publication, it is clear that the epistemic efforts are not extensive since the poll included was not designed to be statistically valid. In the short method section, information about the poll is presented as: "The visualization is not representative for the population as a whole, but shows what these particular persons have responded."

*What is reasonable for us to do?* For DJT, the epistemic efforts in the justifications of truth claims occur within an overarching framework of *reasonableness*. When calculating what epistemic efforts should be made the DJT members weigh up what is practically reasonable (probably a reasoning part of most professions). Reasonableness is traced in the DJT's prioritized work with identifying flaws in data and visualizations. The team makes various control efforts, yet it is evident in our interviews that these efforts mainly involve checking for *visible* errors, that is, what stands out as weird, deviant and so on. A DJT member explains one form of check as follows:

> . . . we can throw in tests that for example, okay we check the last data point that we have, and you know compare it with the next, and then one can see, has it decreased with more than you know, is it down to 0 now (. . .) has it gone from 200 to 0? You know, has it decreased unreasonable much?

Their reasonableness is grounded in their specific context of working conditions, where for example to contact many various instances is seen as difficult to manage in practice:

> But, when it comes to you know the John Hopkins data and stuff like that . . . we don't have the resources so that we can sit and kind of contact them in all the countries in the world you know and verify this data, but one has to stick with the sources that exist. . .

Surely, this focus on the surface (such as numbers that appear as deviant) is sometimes applied in the usage of statistics elsewhere, such as in science, compared to prioritizing practices of checking for potential errors or certain specific interpretations behind numbers or patterns (deviant or otherwise).

## Conclusion

So, by bringing this developing form of independent data journalism into the epistemological investigation through this case study, what is contributed and new about news and its diverse epistemologies? To sum up, this article has advanced the analysis of the epistemology of a distinct form of data journalism labeled here as IDJ, distinct from the use of data in daily news reporting and investigative reporting (Borges-Rey, 2016; Stalph, 2018; Young et al., 2018). IDJ is produced in a context with substantial possibilities for thinking about, and developing, what data journalism can be. IDJ is characterized by the claim of a distinct form of public knowledge (Nielsen, 2017; Park, 1940): a form of public service offering audiences two ways of knowing through visualized data: (1) general information on current topics and (2) opportunities for individualized exploration of data, assuming specific epistemic values for the audiences (Ekström et al., 2022) independent of news values, investigative efforts, or computational logics (e.g. Borges-Rey, 2020; Gravengaard and Rimestad, 2012).

This analysis of IDJ features a multimethod study of knowledge production: what is claimed to be known, the acquiring and processing of information, and the practices and standards applied in justifications of truth claims. IDJ represents a form of knowledge with strong claims to truth by representing "reality" as measurable facts using statistics

and statistics-based visualizations signifying exactness. Numbers, trends and spatial differences constitute the main ways of knowing about the world (Usher, 2020). Simultaneously, they communicate an ambiguous understanding of data, with well-known limitations and uncertainties of data. In the justifications of truth claims, a principle of reasonableness is applied in IDJ to identify the resources that should be prioritized when deciding on epistemic efforts (Ekström et al., 2021; Ettema and Glasser, 1987; Godler and Reich, 2017). Visible errors are in focus while the scrutinizing of the practices of collecting and interpreting data arise less frequently. What is considered reasonable to present as reliable information is decided in relation to what is considered practically feasible to perform.

More broadly, this study ends by discussing two critical questions. First, IDJ raises critical questions with regards the handling of data and their different values for journalism. IDJ as a form of knowledge indicates a tension between producing data visualizations with authoritative "facts" and making use of disclaimers and controlling what is considered reasonable to control. The IDJ epistemology implies an approach to data much in line with an empiricist approach to knowledge. As shown in the analysis, this approach is mirrored in how datasets are both used and checked. Indicated in the team's ambiguity regarding figures as truth, a question for data journalism concerns how to take into account and scrutinize the factors behind the datasets. How data journalists can navigate in this context of providing accurate information using statistics, while not aiming for scientific ambitions, is a challenge for future scholarly interest. Taking into account Parasie (2019) and the promise of transparency, data journalism could enhance people's possibility to understand the elite's calculations about societal issues, yet arguably the factors behind the production of datasets are then relevant objects of journalistic scrutiny.

Second, IDJ is a form of knowledge production providing publicly relevant information through data visualizations. In our study, it is indeed linked to a PSM institution, but we contend that this form of knowledge bears similarities to what other knowledge-producing institutions can achieve. Thus, IDJ expands the boundaries of journalism by leaving news angles behind and using data in its focus on providing general information and individualized data visualizations. In earlier research, Wormer (2017) found science and data journalism to share some characteristics, showing how data journalism is a compromise between journalistic and scientific standards. We argue that IDJ moves further away from traditional journalism by not following traditional news values but does not move in the direction of science as it relies on existing datasets. It does not necessarily produce new knowledge but makes knowledge publicly available. Different ways of producing and providing data journalism assume the forms of knowledge valuable for society. In this way, IDJ adds to news journalism as a form of culture in contributing "assumptions about what matters, what makes sense, what time and place we live in, what range of considerations we should take seriously" (Schudson, 1995: 14).

At last, we stress that IDJ is not possible everywhere in the world, since countries vary in terms of how authorities make datasets publicly available. IDJ gained significance and legitimacy during the COVID-19 pandemic but was also closely linked to working with publicly available pre-justified datasets.

## ORCID iD

Amanda Ramsälv  https://orcid.org/0000-0001-6660-9343

## References

Anderson CW (2018) *Apostles of Certainty*. Oxford: Oxford University Press.

Appelgren E (2018) An illusion of interactivity. *Journalism Practice* 12(3): 308–325.

Appelgren E and Nygren G (2014) Data journalism in Sweden. *Digital Journalism* 2(3): 394–405.

Appelgren E, Lindén C-G and van Dalen A (2019) Data journalism research: studying a maturing field across journalistic cultures, media markets and political environments. *Digital Journalism* 7(9): 1191–1199.

Borges-Rey E (2016) Unravelling data journalism. *Journalism Practice* 10(7): 833–843.

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.

Cheruiyot D, Baack S and Ferrer-Conill R (2019) Data journalism beyond legacy media: the case of African and European civic technology organizations. *Digital Journalism* 7(9): 1215–1229.

Coddington M (2015) Clarifying journalism's quantitative turn. *Digital Journalism* 3(3): 331–348.

Cushion S, Lewis J and Callaghan R (2017) Data journalism, impartiality and statistical claims. *Journalism Practice* 11(10): 1198–1215.

de Maeyer J, Libert M, Domingo D, et al. (2015) Waiting for data journalism. *Digital Journalism* 3(3): 432–446.

de-Lima-Santos MF and Mesquita L (2021) Data journalism in favela: made by, for, and about forgotten and marginalized communities. *Journalism Practice*. Epub ahead of print 5 May. DOI: 10.1080/17512786.2021.1922301.

Desai A, Nouvellet P, Bhatia S, et al. (2021) Data journalism and the COVID-19 pandemic: opportunities and challenges. *The Lancet Digital Health* 3(10): E619–E621.

Ekström M (2002) Epistemologies of TV Journalism: a theoretical framework. *Journalism* 3(3): 259–282.

Ekström M and Westlund O (2019) Epistemology and journalism. In: *Oxford Encyclopedia of Journalism Studies*. Oxford: Oxford University Press.

Ekström M, Ramsälv A and Westlund O (2021) The epistemologies of breaking news. *Journalism Studies* 22(2): 174–192.

Ekström M, Ramsälv A and Westlund O (2022) Data-driven news work culture: Reconciling tensions in epistemic values and practices in news journalism. *Journalism* 23(4): 755–772.

Engebretsen M, Kennedy H and Weber W (2018) Data visualization in Scandinavian newsrooms: emerging trends in journalistic visualization practices. *Nordicom Review* 39(2): 3–18.

Ettema JS and Glasser TL (1987) On the epistemology of investigative journalism. In: Gurevtich MR and Levy M (eds) *Mass Communication Yearbook*. London: SAGE, pp. 338–361

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

Godler Y and Reich Z (2017) Journalistic evidence: cross-verification as a constituent of mediated knowledge. *Journalism* 18(5): 558–574.

Gravengaard G and Rimestad L (2012) Elimination of ideas and professional socialisation. *Journalism Practice* 6(4): 465–481.

Helberger N (2019) On the democratic role of news recommenders. *Digital Journalism* 7(8): 993–1012.

Heravi B, Cassidy K, Davis E, et al. (2022) Preserving data journalism: a systematic literature review. *Journalism Practice* 16: 2083–2105.

Jacobson S, Marino J and Gutsche RE Jr (2015) The digital animation of literary journalism. *Journalism* 17(4): 527–546.

Karlsson M (2010) Rituals of transparency. *Journalism Studies* 11(4): 535–545.

Kligler-Vilenchik N and Tenenboim O (2020) Sustained journalist–audience reciprocity in a meso news-space: the case of a journalistic WhatsApp group. *New Media & Society* 22(2): 264–282.

Kovach B and Rosenstiel T (2001) *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. 1st ed. New York: Three Rivers Press.

Lawson BT (2021) Hiding behind databases, institutions and actors: how journalists use statistics in reporting humanitarian crises. *Journalism Practice*. Epub ahead of print 27 May. DOI: 10.1080/17512786.2021.1930106.

Lewis NP and Al Nashmi E (2019) Data journalism in the Arab region: role conflict exposed. *Digital Journalism* 7(9): 1200–1214.

Lewis NP and Waters S (2018) Data journalism and the challenge of shoe-leather epistemologies. *Digital Journalism* 6(6): 719–736.

Loosen W, Reimer J and 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.

Lugo-Ocando J and Nguyen A (2017) *Developing News—global Journalism and the Coverage of "Third World" Development*. London: Routledge.

Matheson D and Wahl-Jorgensen K (2020) The epistemology of live blogging. *New Media & Society* 22(2): 300–316.

Mutsvairo B, Bebawi S and Borges-Rey E (2019) *Data Journalism in the Global South* (ed B Mutsvairo, S Bebawi and E Borges-Rey). Cham: Springer International Publishing.

Nielsen R (2017) Digital news as forms of knowledge. In: Bozkowski PJ and Anderson CW (eds) *Remaking the News*. Cambridge, MA: The MIT Press, pp. 99–110.

Ojo A and Heravi B (2018) Patterns in award winning data storytelling. *Digital Journalism* 6(6): 693–718.

Palomo B, Teruel L and Blanco-Castilla E (2019) Data journalism projects based on user-generated content. How la nacion data transforms active audience into staff. *Digital Journalism* 7(9): 1270–1288.

Parasie S (2015) Data-driven revelation? *Digital Journalism* 3(3): 364–380.

Parasie S (2019) Data journalism and the promise of transparency. In: Tumber H and Waisbord S (eds) *The Routledge Companion to Media and Scandal*. 1st ed. London: Routledge, pp. 263–272.

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.

Park RE (1940) News as a form of knowledge: a chapter in the sociology of knowledge. *American Journal of Sociology* 45(5): 669–686.

Pentzold C, Fechner DJ and Zuber C (2021) "Flatten the curve": data-driven projections and the journalistic brokering of knowledge during the COVID-19 crisis. *Digital Journalism* 9(9): 1367–1390.

Reich Z (2011) Source credibility and journalism. *Journalism Practice* 5(1): 51–67.

Royal C and Blasingame D (2015) Data journalism: an explication. *International Symposium in Online Journalism* 5(1): 24–46.

Schudson M (1995) *The Power of News*. Cambridge, MA: Harvard University Press.

Schultz I (2007) The journalistic gut feeling. *Journalism Practice* 1(2): 190–207.

Splendore S (2016) Quantitatively oriented forms of journalism and their epistemology. *Sociology Compass* 10(5): 343–352.

Stalph F (2018) Classifying data journalism. *Journalism Practice* 12(10): 1332–1350.

Tandoc EC Jr and Oh S-K (2017) Small departures, big continuities? *Journalism Studies* 18(8): 997–1015.

Usher N (2018) Breaking news production processes in US metropolitan newspapers: immediacy and journalistic authority. *Journalism* 19(1): 21–36.

Usher N (2020) News cartography and epistemic authority in the era of big data: journalists as map-makers, map-users, and map-subjects. *New Media & Society* 22(2): 247–263.

Widholm A and Appelgren E (2022) A softer kind of hard news? Data journalism and the digital renewal of public service news in Sweden. *New Media & Society* 24(6): 1363–1381.

Wormer H (2017) Mind the statistics gap: science journalism as a bridge between data and journalism. In: Nguyen A (ed.) *News, Numbers and Public Opinion in a Data-driven World*. 1st ed. New York: Bloomsbury Academic, pp. 226–241.

Wright K, Zamith R and Bebawi S (2019) Data journalism beyond majority world countries: challenges and opportunities. *Digital Journalism* 79(9): 1295–1302.

Wu S (2021) Data "objectivity" in a time of coronavirus: uncovering the potential impact of state influence on the production of data-driven news. *Digital Journalism* 9(9): 1303–1320.

Young ML, Hermida A and Fulda J (2018) What makes for great data journalism? *Journalism Practice* 12(1): 115–135.

Zamith R (2019) Transparency, interactivity, diversity, and information provenance in everyday data journalism. *Digital Journalism* 7(4): 470–489.

## Author biographies

Amanda Ramsälv is a PhD candidate at the Department of Journalism, Media and Communication, University of Gothenburg. She has a background in the discipline of Philosophy, and working as a journalist. Her dissertation focuses on journalism as a central knowledge-producing institution.

Mats Ekström (PhD) is a professor at the Department of Journalism, Media and Communication, University of Gothenburg. His research focuses on journalism, political communication, media discourse, and conversations in institutional settings.

Oscar Westlund (PhD) is professor at Oslo Metropolitan University, where he co-leads the OsloMet Digital Journalism Research Group. He holds a secondary appointment at the University of Gothenburg and is the Editor-in-Chief of Digital Journalism. He specializes in digital journalism, fact-checking, platforms, media management, news consumption, and mobile media.