Studies in Communication Sciences 18.1 (2018), pp. 191–206

Data stories. Rethinking journalistic storytelling in the context of data journalism

Wibke Weber, ZHAW Zurich University of Applied Sciences, School of Applied Linguistics*
Martin Engebretsen, University of Agder, Department of Nordic and Media Studies
Helen Kennedy, University of Sheffield, Department of Sociological Studies
*Corresponding author: wibke.weber@zhaw.ch

Abstract
This paper addresses the increased use of data and data visualization in newsrooms, which has yielded a new form of storytelling: data stories. In journalism, data stories or storytelling with data are the new buzzwords. What journalists mean by data stories, however, remains blurred. We use the emergence of data stories as an opportunity to describe the changing understanding of journalistic storytelling. Based on interviews with editorial leaders, data journalists, developers, and designers in 26 major news organizations in Europe, we focus on practitioners' perspective on data stories. In our empirical study, we identified seven key features of journalistic data stories: data, communicative function, the textual-visual relationship, structure and design of a story, interactivity, and the meta-story. These findings contribute to rethinking the narrative approach to journalism.

Keywords
data stories, data visualization, narrativity, storytelling, data journalism, newsroom

1 Introduction

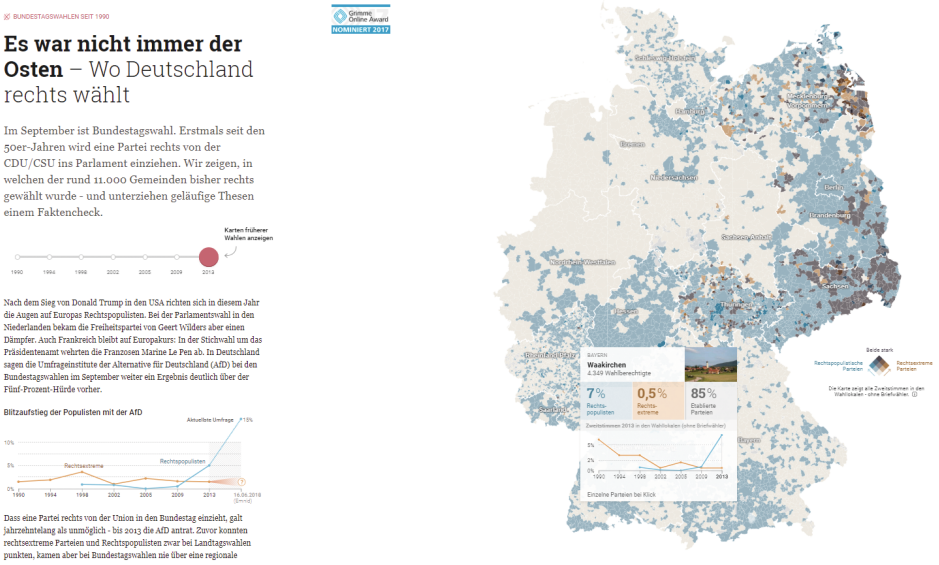
The way journalists understand and do journalism is changing. Digital technologies, big and open data, and the datafication of many aspects of life are permeating newsrooms and have led to new forms in journalism. One of these forms is data journalism. Data journalism affects professional practices in newsrooms and thus the artifacts produced within this professional domain. Today, "(j)ournalists, data scientists and computational journalists are all storytellers" (Marconi, Siegman & Machine Journalist, 2017, p. 6). This statement is indicative of changes in journalistic storytelling: from writing as the main semiotic mode to coding and visualizing as pivotal elements of digital storytelling.

One key definition of data journalism, often cited in the relevant literature, blogs, and websites is that data journalism involves "gathering, cleaning, organizing, analyzing, visualizing and publishing data to support the creation of acts of journalism" (Howard in Hamilton, 2016, p. 297; Sunne, 2016). Rinsdorf and Boers (2016, p. 1) define data journalism as "a qualitatively new way of reporting which gains insights about relevant societal trends by analyzing open datasets using (semi-)automatized methods to detect meaningful patterns in data structure." In this sense, data journalism is understood as both a process, i. e., analyzing large datasets and telling stories with data, and a product, including data visualization and other textual outputs based on data (Ausserhofer, 2017, p. 4). Arguably, data journalism as a process has the ability to reveal the story within data, and data visualization as a product – a visual representation – enables its users to see this story.

With the emergence of data journalism, there is also an increased use of data visualization in newsrooms (Rogers, Schwabish, & Bowers, 2017), and data visualizations are at the heart of data stories. Usually, data visualization plays a significant role in data stories (although it is also possible to build data stories on a purely textual and numeric basis). Like info-

https://doi.org/10.24434/j.scoms.2018.01.013

Figure 1: Interactive map about "It wasn't always the East – where Germany votes for the far-right" combined with a timeline slider and text (courtesy of Berliner Morgenpost)[a]

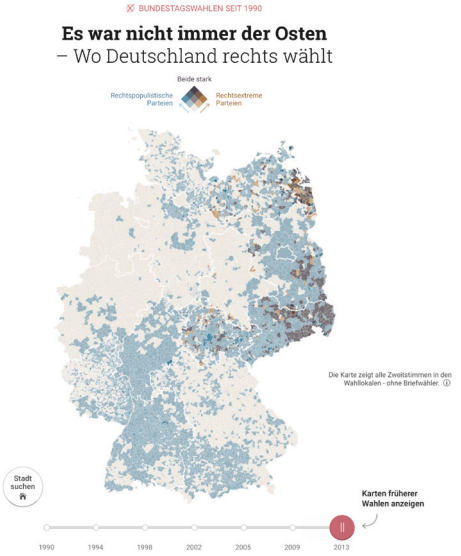


[a] Berliner Morgenpost (23.01.2017), https://interaktiv.morgenpost.de/wo-deutschland-rechts-waehlt/

graphics, data stories can be described as multimodal hybrid artifacts that weave together numbers, words, images and design into a coherent whole (Weber, 2017; Engebretsen & Weber, 2017; Kennedy, Hill, Aiello, & Allen, 2016; Cairo, 2016, 2013; Weber & Rall, 2016). The close relatedness of data visualization and data story might explain why these terms are often used synonymously by practitioners.

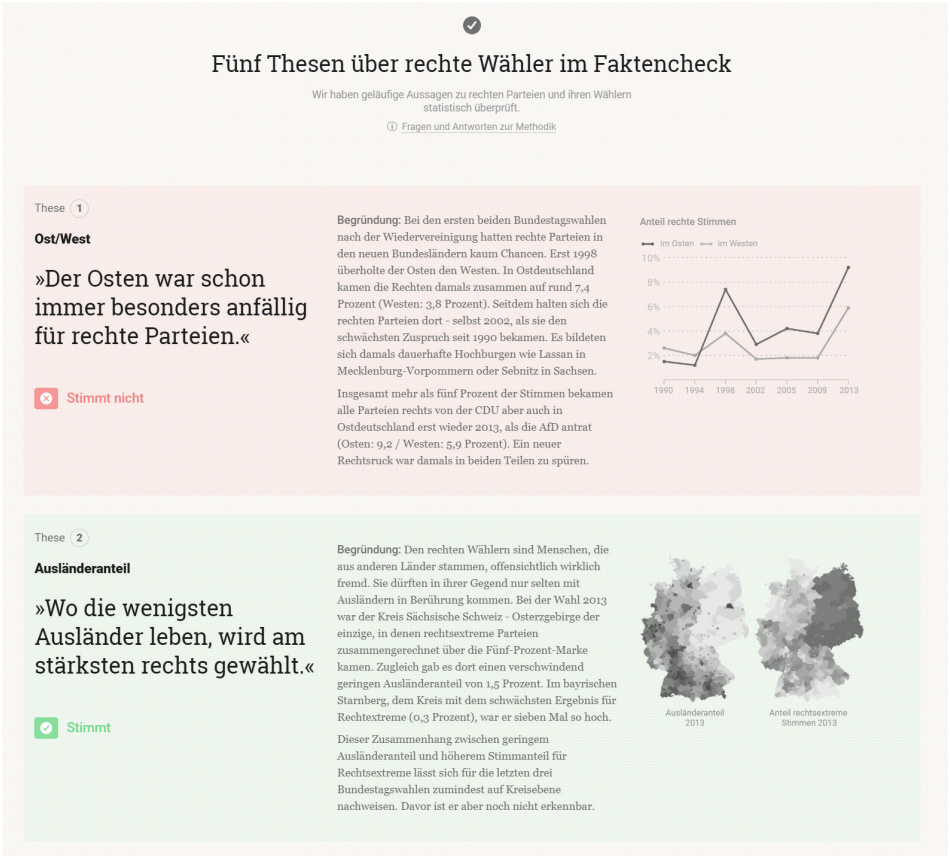
In this article, we focus on data stories based on data visualizations that range from simple bar charts to complex multimodal and interactive stand-alone graphics. The article is informed by a qualitative interview study on uses of data visualization in European newsrooms. The study included 60 interviews conducted in 26 newsrooms (see Section 3). In the study, we investigated the extent to which data visualization represents a significant change in the practices of journalistic storytelling. The overarching questions of this article are: What are the key features of journalistic data stories as seen from the newsroom and to what extent does the emergence of data stories require us to rethink under-

Figure 2: Mobile version of the interactive map "It wasn't always the East – where Germany votes for the far-right" combined with timeline slider and search option (courtesy of Berliner Morgenpost)[a]



[a] Berliner Morgenpost (23.01.2017), https://interaktiv.morgenpost.de/wo-deutschland-rechts-waehlt/

Figure 3: Fact check boxes with a claim (left side), the written argument (middle) and data visualizations (right side) providing the facts. Excerpt taken from "It wasn't always the East – where Germany votes for the far-right" (courtesy of Berliner Morgenpost)[a]



[a] Berliner Morgenpost (23.01.2017), https://interaktiv.morgenpost.de/wo-deutschland-rechts-waehlt/

standings of journalistic storytelling? The aim of this paper is threefold: (i) to develop understanding of the hitherto vague term "data stories" by merging theoretical and empirical perspectives on journalistic storytelling, (ii) to identify key features that characterize "storytelling with data," and (iii) to contribute to the research field of journalistic storytelling through our focus on data visualization by showing that a narrative approach does not necessarily fall short, but that it needs revision.

Figure 1 illustrates what we mean by data story. The multimodal feature titled "It wasn't always the East – where Germany votes for the far-right" is about the development of right-wing populism, far-right parties and extreme right parties in Germany between 1990 and 2013 and was published 2017 by the Berliner Morgenpost, a regional daily newspaper.

The core of the feature is an interactive choropleth map linked to a time slider to reveal changes in voting behavior over the years. Readers can also browse through the map looking for cities they are interested in. The data visualization can be regarded as a stand-alone graphic (a story map) that contains a narrative in itself: a beginning, a change, and an ending. At the same time, the data visualization works as the starting point for the multimodal feature. As such, it appears at the top of the web page and the mobile page (Fig. 2). The whole feature

consists of a video, texts, further graphs with annotations, five fact check boxes (Fig. 3), and a clickable icon that links to the data collection and analysis methodology, what we call the meta-story.

In the following sections, we first outline theoretical approaches and studies related to our research focus (Section 2). This is followed by a description of the methodology applied in our study (Section 3). Then we present the findings: the key features of data stories that lead to rethinking the narrative approach in journalism (Section 4). The article concludes with a summary of the main findings and an outlook on further research questions (Section 5).

2 Related work and theoretical approaches

As a relatively new phenomenon the term "data stories" is widely used on blogs, websites, and podcasts (e. g., Bertini & Stefaner, n. d.). However, seen from a scholarly perspective, it lacks a clear definition. Data stories are based on *numerical and/or categorical data*[1]. So, the starting point of the journalistic work is collecting data or searching for open datasets and analyzing them with the aim of finding patterns, clusters or statistical outliers that are newsworthy and worth sharing. The findings are mostly presented in data visualizations. Based on several definitions, we define data visualization as a visual representation of data created to amplify the cognitive processing and the social application of the data represented (Borgo et al., 2013; Card, Mackinlay, & Shneiderman, 1999). Classic data visualizations are graphs, charts, maps and timelines, or a combination of these.

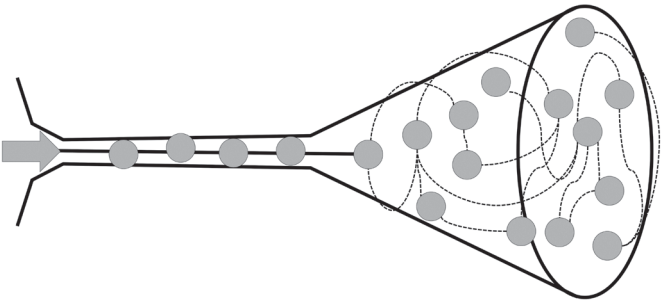
The term "data stories" suggests that we are dealing with narrative. A narrative is defined by Alan Bell (1991) as a sequence of events that are temporally structured and coherently related to each other with bonds of (strong or weak) causality. Jahn (2017) defines story in a similar way, as "a sequence of events involving characters"; he further describes events as "natural and non-natural happenings." An example of a minimal story given by E. M. Forster (1974, p. 93 f.) is "The king died and then the queen died." For that we need a beginning, an end, and a change in between. However, we know from ethnographic studies in the field of newsmaking that journalists often use "story" and "news" synonymously because both refer to happenings (Merminod, 2016; Perrin, 2015). When journalists talk about storytelling, they mean not only *"narrating"* but also other text-linguistic practices such as *describing, explaining,* and *arguing* (Perrin, 2015, 2011). These text-linguistic practices have been described by Brinker (2010), who distinguishes between four basic ways in which a topic can be presented: descriptive, narrative, explicative, and argumentative. Thus the distinction researchers make theoretically between describing, narrating, explaining, and arguing remains blurred in the field of practice.[2]

Similar terminological slippage can be found in a report on data journalism by Rogers, Schwabish and Bowers (2017). Based on a survey, they provide a classification of data stories that is "based on the types of news produced from it" (Rogers, Schwabish, & Bowers, 2017, p. 6). The three main types are: "(i) stories that are enriched by data; (ii) stories that use data to investigate; (iii) stories that explain data." The classification and the report's use of the term "story" address the work of data journalists in general, rather than understanding stories as defined in the field of narratology. That is why, seen from the theoretical perspective of narratology, journalistic storytelling often appears fragmentary. In this regard, Kosara (2017) states that in the field of data visualization, many so-called data stories only present facts without a story arc. In his case study, he describes different narrative patterns and argument structures in data stories. Segel and Heer (2010) also identified dif-

1 The concepts and techniques relevant to the coding process are highlighted in italics.

2 Describing means showing by giving answers to the four Ws "who," "what," "when," "where." Explaining and arguing addresses questions concerning the "how" and "why."

Figure 4: Narrative pattern of the linear-nonlinear type: the Martini glass structure. (Source: W. Weber)



ferent narrative structures in data stories. They distinguish between an "author-driven" and a "reader-driven" approach by connecting two inherent components that characterize data stories: *narrativity and interactivity.* According to Ryan (2006, p. 99), a *linear structure* enables the producers to tell a story ("top-down-design"). This is what Segel and Heer (2010, p. 1146) call the *"author-driven"* approach, which fits best "when the goal is storytelling or efficient communication." In contrast, a highly interactive *nonlinear dramaturgy* requires a "bottom-up-input from the user" (Ryan, 2006, p. 99). Segel and Heer call this nonlinear type the *"reader-driven approach"* (Segel & Heer, 2010, p. 1146). In this case, the user is given maximum information to explore and the data visualization turns into "a tool for readers to analyze what's being presented to them" (Cairo, 2013, p. 73).

Experts and practitioners in the field of data visualization and data journalism often refer to the terms "reader-driven" (or user-driven) and "author-driven" as *exploratory* and *explanatory* (Bradshaw, 2017; González Veira, 2017). The journalist Paul Bradshaw discusses this point on his blog: "At the heart of this move from the exploratory to the explanatory is a problem that game designers have grappled with for years: how much do you let someone explore a world (of information, in this case), and how much do you exercise editorial control in order to create an experience that larger groups are going to want to engage with?" He continues, "How much responsibility to [sic] we have for the stories that people tell with our information? And how much responsibility do we have for delivering as much information as someone needs? This is the story vs information problem in a nutshell" (Bradshaw, 2017).

A third category of data story exists that results from the hybridization process of the two types: the *linear-nonlinear type* (Weber, 2017, 2013). The main story line is given by the author, but the user may explore datasets at a certain stage of the story. The advantage for users is that the information is conveyed in a structured way with the option to explore the data to a certain degree. A variation of this linear-nonlinear type is the *"Martini glass structure"* (Segel & Heer, 2010), where the users are guided step by step through the visualization and at the end they can explore the data by themselves (Fig. 4).

Further research has recently been conducted to advance the discussion on narrative visualizations. Brehmer, Lee, Bach, Riche, and Munzner (2016) found that timelines and time series, which offer a sequence of narrative points, and visualization sequences (e.g., small multiples) seem to work better to transform data into stories (according to Bell's definition) than simple charts or exploratory data visualizations (Hullman, Kosara, & Lam, 2017; Brehmer et al., 2016). According to Brehmer et al., smooth animated transitions between narrative points can foster the co-

herence of a story. Stolper, Lee, Riche, and Stasko (2016) looked for narrative techniques that data-storytellers use today and compared them to those identified by Segel and Heer (2010). Informed by an analysis of a large dataset of visualizations with a focus on asynchronous, author-driven data stories, Stolper et al. present several new storytelling techniques such as *linking between text and chart,* creatively constraining the reader through scrolling, brushing and linking between visualizations, linking elements through color or animation, or the *scrollytelling technique*. The simplest narrative technique they found is using a textual narrative and interspersing visualizations throughout. Timelines, slide shows, and data videos are often used to trigger storytelling because of their inherent linear structure. Scrollytelling involves a story unfolding as the user scrolls down the page and the visualizations that are embedded in the article appear at the appropriate time. Scrolling also triggers changes in the visualization itself, e. g., zooming out from the visualization (Stolper et al., 2016). In this case, both the visualization and text form the narrative.

From these theoretical considerations and empirical studies, we are able to deduce criteria for our analysis of our interviews. The methodology we used to produce these findings will be explained in the following section, which is followed by a presentation of the findings themselves.

3 Methodology

To understand practitioners' perspectives on storytelling with data we interviewed 60 data journalists, designers, developers, and newsroom leaders in 26 major news organizations in six European countries: Norway (NO), Sweden (SE), Denmark (DK), Germany (DE), Switzerland (CH), and the United Kingdom (UK). The interviews took place from March 2016 to February 2017. The newsroom types range from national broadcasters, national broadsheet and tabloid newspapers to regional broadcasters and newspapers and online news providers. The interviews were conducted face-to-face in the respective newsroom or via video conferencing according to a semi-structured interview guide. Each interview took about one hour, was recorded as audio file and then transcribed and anonymized. To aid comparison, the main aspects of the Scandinavian and German-speaking interviews were translated into English.

The results presented here are one of several outputs of the interview study (see also Engebretsen, Kennedy, & Weber, 2017). For the purposes of this paper, the focus of our analysis lies on journalistic storytelling and narrative data visualization. For the qualitative content analysis of the data corpus (Mayring, 2000), we used the software tools MAXQDA and Nvivo. Five coding categories were developed deductively based on the literature review in section 2: *data, communicative function, textual-visual relationship, structure* or *architecture,* and *interactivity*. Two further categories emerged inductively from the interview data: *design* and *meta-story*. By design we mean – in a very simplified definition – the dimension of the visual performance. By meta-story, we mean text elements produced in order to make the journalistic methods transparent which is regarded as crucial in data journalism (Rinsdorf & Boers, 2016; Matzat, 2016), such as a "How-we-did-it"-story.

We coded our interviews by looking for these features:

1 *Data:* statements regarding how the respondents deal with data as a basic element of data stories.
2 *Communicative function (purpose):* key words indicating the communicative intention of the producers, e. g., to tell, to explain, to argue visually.
3 *Textual-visual relationship:* the interrelationship or interplay between text and visualization.
4 *Visual Design:* keywords like aesthetics, style, performance, looking good, attractive.
5 *The structure or architecture:* linear and nonlinear ways of storytelling.

6 *Interactivity:* users' involvement and engagement with the story and the option to interact with the data visualization (e.g., "search for," "find out," "explore").
7 *The meta-story:* transparency concerning data sources and data analysis.

In what follows, we summarize the findings from our analysis of the interviews.

4 Results

The coding process has led to seven key features that characterize journalistic data stories. These key features are corroborated by anonymized excerpts taken from the interviews. The origin of each citation is indicated by the respective country in parentheses after the quote to underline that this is not a single statement or phenomenon in only one news organization but a trend we have identified in several newsrooms across different countries.

4.1 Data: the core of data storytelling

Our study clearly indicates that data is increasingly used to find a story and often shapes the core of a story. As mentioned above, story is defined as a coherent sequence of events. In data journalism, data teams first have to find these "events" in the data by data mining, scraping websites, filtering and analyzing datasets. Many interviewees stressed the possibility of finding new stories that could not be told without the analysis and visualization of data.

> Sometimes the data is almost a story and sometimes the data supports a story. (…) All the stories they do have the data almost as a starting point; the data is really interesting and from there they reveal things or they find interesting things, so in that case the data is the start of the story. (UK)

> And so [text-centric journalism] is what we're trying to fight against now, is the idea that actually if you do data and graphics properly, (…) it might help you to find the stories in the first place, and people like (ANONYMIZED) here are great data journalists who are actually finding stories by deeply diving into data in the first place. (UK)

Numbers and statistics are not completely new in journalism, but new software tools and the opening up of public data help journalists to use them more analytically. Now, data journalists consider data as an essential semiotic resource in news reporting, similar to verbal texts, images and audio.

> Someone described it quite nicely as it being like a lump of clay and you've got to get a good feel for it before you decide whether it's going to be a bowl or a vase. (…) I definitely would say that with numbers, it's good to explore them and turn them on their heads and try them in different ways and see where the story is. (UK)

> An important aspect of modernizing our journalism is to dig into large databases, understand what they tell us and present it visually for the reader. (NO)

A question, a problem, or an idea can precede the data search.

> At the beginning, you have an idea or a dataset or a data source from which the story emerges. Both are possible. If the idea is at the beginning, it takes the most effort to search for data; it often takes weeks or months for data collecting. If you already have the dataset at the beginning, then it's easier. (DE)

One point that is striking is that coding from scratch can be seen as a new element that forms the story and a highly valued skill for journalistic storytelling (cf. Rogers et al., 2017, p. 9).

> We, the interactive team, set up our own system. That means, we are right in the code. (DE)

4.2 The communicative function: the act of showing

Data stories based on data visualization show something visually that is hard to

explain verbally. They intend to make the reader see and facilitate understanding (Kirk, 2016, p. 21). Our analysis of the interviews has shown that data teams employ narrative, explanatory, and argumentative techniques, even though they do not use such terms.

> Saying, look, here you can see the way in which something has developed – this works as an argument. (CH)

> I believe that this is the core of our mission, to try to mediate what is correct and true. To find the facts. And facts are often represented by numbers. How many inhabitants are there in our city? That number is a fact, and it does not lie. (NO)

> We're trying to say look, we've looked at this, it's difficult data but we've done that and now we can communicate quite a simple story to you. (UK)

Common terms interviewees used to describe their communicative intentions are "to offer insight" (UK), "to explain more easily" (SE), "to communicate clearly, more clearly than words can" (UK), "to tell several facets in detail, which in text is only possible in an aggregated form" (DE), to make stories "more accessible" (DK), "to reveal deplorable state of affairs" (CH), "to help people understand the world" (UK).

These statements corroborate the assertion of Rinsdorf and Boers that new practices of data journalism may be seen as a "part of a general switch in journalism from a focus on news and scoops to background information and the explanation of current trends" (Rinsdorf & Boers, 2016, p. 2). Thus, data stories can work as background stories in order to give more detailed explanations.

Moreover, data visualization is used to argue visually by providing empirical evidence. Thus, it gives support to the claims made verbally in the text (Kosara, 2017).

> It's when the work you're doing explains things that are perceived in different ways, so it's when you are trying to communicate something new, something that is hidden behind the data, something that most of the people perceive in different ways, then you feel like you're really trying to communicate truth. (UK)

> Visualizations help to convey things in a better way, to interpret things and to learn things. Visualizations are tools, visual arguments, and in this regard they do better than text. (DE)

In data stories, with large amounts of visual-verbal content, the scrollytelling technique is often employed for the interplay between visualization and text, which will be discussed further in the next section.

4.3 The textual-visual relationship

Many interviewees see potential for new narrative techniques through their use and visualization of data. Two main patterns of text-visualization relationship can be identified from the interviews. The first is data visualization as a stand-alone graphic, which represents a single multimodal unit and coherent whole or a story in itself, usually followed by an article that provides additional information and further aspects. This kind of visualization is often placed at the top of the page as an eye-catcher and starting point of a longer feature or long-form article. In this way, one interviewee argues that the data visualization is more than an add-on; it plays a role in defining the structure of the story:

> So, in fact it would be completely wrong to do the graphics at the end of the process because, if you do them well enough, they should be influencing the structure of your text. So, we're on a journey towards that. (UK)

The second pattern is text alternating with data visualization and other elements (e. g., video, audio) in a sequential structure.

> A new trend is that we divide an article up in distinct sections, instead of just a wall of text. And a graph can here work as a nice break in the reading, and it can be placed exactly where it belongs in the story. (NO)

> It is easy to insert the visualization into the text – exactly at this place where the text talks about the issue. So, we cut the print graphic in several pieces and scatter it over the online text. (CH)

> We are now developing a new template, more flexible than the standard news template. (…) Here we can insert images, citations, diagrams, videos etc. in addition to text elements. I think it will inspire us to think in new ways about news stories. (SE)

Many interviewees stated that data visualization is the driving force of a data story, even when it is a simple graphic or diagram. According to click-through rates, visual data stories are more attractive than text-based stories.

> The editor-in-chief of Al Jazeera emphasizes: give the people broccoli and ice cream. That's their principle. Ice cream is the lead, the teaser, that is the visualization; and broccoli is the hard stuff below, for those who are interested in it, but if you want you can skip it. (DE)

> The reader stats tell us that when we insert a simple data visualization in a story, readers stay on the page a little longer. (SE)

> So, the copy is written – so we decide together on the structure and on the story itself, and then the story flows and the visuals are really integrated in the story. So, they are part of the story. They are the story. And the copy is integrated as well. So, you know, it's visuals and text. (UK)

> Often in our stories we use the scrolling technique. It is not necessary to click but to scroll, if you scroll down, something will happen in the story. That's how we tell the story, how we do it – maybe the correct term for this is scrollytelling. (DE)

However, to keep readers interested and hooked, it is not enough to develop a visual data story. Most interviewees acknowledged that the story necessarily has to look good. As we will outline in the next section, an attractive and engaging design therefore plays a central role in data storytelling.

4.4 Visual design matters

In the context of a data story, design concerns all the visible elements that create the visual performance of the data story: *how* the data is presented. Visual design plays a major role in data stories and in multimodal artifacts in general (e. g., van Leeuwen, 2005; Kress & van Leeuwen, 1996). Design elements such as font, color, size, alignment, form, position, texture, lines and spacing work as organizers, providing orientation, cohesion, and coherence to a journalistic artifact (Engebretsen & Weber, 2017; Weber & Rall, 2016; Kennedy et al., 2016; Hullman & Diakopoulos, 2011).

There was a near consensus among the interviewees that aesthetics is more than "nice to have."

> It is getting more and more important. Nobody looks at a graphic that looks bad. (DE)

> When I read a story that is well designed and embedded in an aesthetic context, I consider the story more reliable. (DE)

> Obviously we've got to get the informative done after we engage them, so they've got to look pretty to engage. (UK)

> You have a much better chance to attract the younger readers if you present them with an image and some kind of dynamics. Maybe it is more aesthetic than it is actually pedagogical, but it might still be a good thing to do, because you reach a group not very interested in news. (SE)

> And sometimes our readers will use some more time on the story because the graphic looks good. (DK)

> Simply because aesthetics, you know, they help you get the story in a better way. So, you know, if it looks nicer, then you will spend more time with it, basically. (UK)

According to the interviewees, an attractive and user-friendly design is a prerequisite for capturing the readers' interest

and drawing them into the story. Then, the structure of the story is the crucial point in establishing whether the data visualization is engaging or not.

4.5 The structure: tending towards the Martini glass structure

As mentioned in Section 2, there are three main patterns for developing data stories: (i) the linear, author-driven approach with a step by step dramaturgy; (ii) the non-linear exploratory reader-driven approach; (iii) the hybrid linear-nonlinear pattern known as the Martini glass structure. It emerged from the interviews that some data teams are tending towards the Martini glass structure, first to tell the basic story in a linear way and then to open up the data visualization for exploration.

> There's that old Martini glass theory about data visualisation, which is: give somebody something to hold onto and then hit them with the broader stuff later, and I think there is something in that, certainly this is one of our most successful ways of telling stories. (UK)

> And you could offer the content step by step – first show one sample in the news story and then say: please, you are now welcome to explore the rest of the data! (NO)

> It depends on the story we want to tell and it depends on how much we expect the reader to spend because most of the time readers don't have the time to play with the stuff we produce, so we have to find the right balance and to produce some work that helps the reader to understand at the first glance but also gives the possibility to dig more into the content so in that way we cover both. (UK)

One interviewee outlined that sometimes they offer a series of Martini glass-structured data visualizations.

> We used a structural principle I found works well. It was linear on the story-level, but with numerous openings for further exploration at different points in the main story. So, it was both linear and networked. (NO)

While the stem of the Martini glass – the step by step pattern – symbolizes a sequential structure, the bowl enables users to interact with the graphic, which leads us to the next key feature: interactivity.

4.6 Interactivity: find *your* story

The main difference between text-based stories and visual data stories with non-linear exploratory elements is that the latter comprise potentially manifold narratives embedded in one single data visualization. They possess the ability to foster engagement, i. e., to encourage users to look for further stories on their own. Thus, interactive data stories create a modern form of dialogic journalism (cf. Engebretsen, 2006). One interviewee called it "You-journalism," and many confirmed this point, as seen in the following quotes:

> We sometimes have a large dataset underlying our news stories, and sometimes we give them an interactive interface for the reader to explore by herself. That creates a modern form of "You-journalism," as it gives you the opportunity to check out your own neighborhood or something similar. In that way, we can offer much more information than just the one example we select to angle the news story. (NO)

> We found the best way to connect with the audience and get them to really engage and understand the story is to make it personally relevant to each individual one of them. (UK)

To attract users' interest, data teams usually demonstrate in the basic story how to use the data visualization ("this is what the data can tell you here") before leaving the data to the users. Thus, they expect a higher level of interest than in text-based stories.

> When we say to the readers "Here you can check out your own local area," it means that the story becomes stronger than when we pick out one case to illustrate the general topic. (NO)

> So sometimes we put the data out there in the visual, but in a more exploratory way. So, peo-

> ple can look for their own – you know, if it's geography, they can search by their location and they can get the micro stories. They can get into the stories through their location and through their own interests. (UK)

> This could be a map, which tells one possible story very concisely, and at the end the reader has the option to look for his or her own location, how does it affect me. But the basic story is already told. (DE)

Sometimes this kind of data story is built on a quiz format to increase engagement.

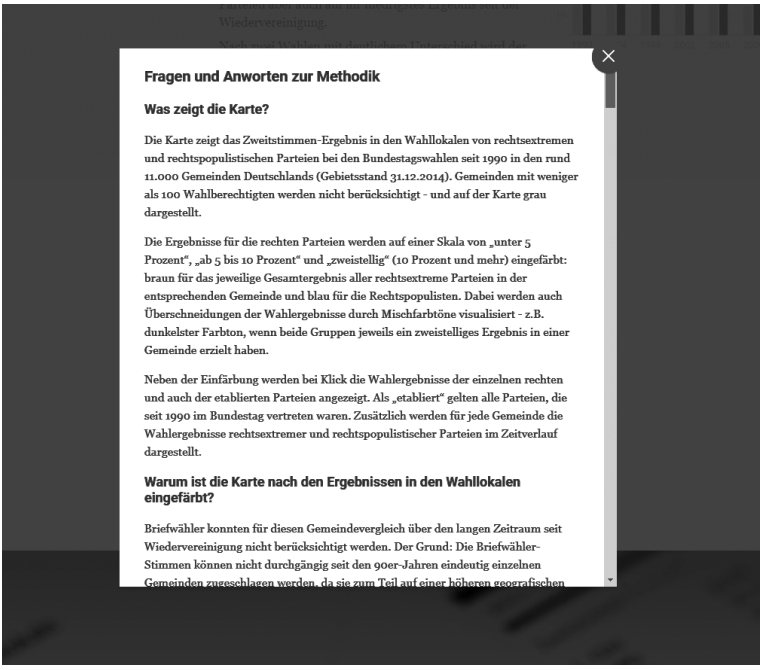
> In the last months, we try some new things, not very big and not the classical visualizations, but interactive, such as quizzes. We think they are a good way because they are playful and they are well suited to conveying complex content. (CH)

To communicate their journalistic practices clearly, data stories are usually linked to a meta-story, which will be taken up in the next section.

4.7 The meta-story: making processes transparent

The process of producing data visualizations is a more complex and non-transparent one than producing verbal texts and photos – and thus calls for a stronger awareness of transparency and trust-building. In data journalism, transparency has become a common principle (Sunne, 2016). For most of the interviewees, transparency is a sine qua non. Transparency means: stating the source, including a link to the data sources, providing access to the raw data (e.g., downloadable as Excel files or on platforms like Github), explaining the methodology, or all of these (Fig. 5).

Figure 5: Questions and answers concerning the methodology of "It wasn't always the East – where Germany votes for the far-right." In addition, a link is provided to download the data (courtesy of Berliner Morgenpost)[a]



[a] Berliner Morgenpost (23.01.2017), from https://interaktiv.morgenpost.de/wo-deutschland-rechts-waehlt/

> It is important to include verbal explanations and we say what has been omitted and for what reason. Sometimes we publish an extra article to explain: what can data tell us and what not. (CH)

> It's important to explain how and why you decided to go that way so you make things clear and you explain to the reader there is also another side of the story, like being clear by explaining methodology, by giving the proper scales, by giving the proper legends, annotations and data sources, all of them together will help the reader understand, and reassure actually the reader also that what you're showing is not something built on just some perception. (UK)

Interviewees regarded the meta-story as an important part of the data story. This "how-we-did-it"-element can be integrated into the story, but usually it appears at the end of a story in a fact box or as a link (Figure 5).

> We need to make room for an explanation, typically a fact box. Sometimes the issues we are working on can be so controversial that we need to make a separate story about what we have done with the data. (DK)

> Sometimes, this is a part of the story: to explain the methods applied. There are also cases where the methodology must be described very clearly and in detail, so it is not possible to incorporate the methods in the story because it would take up too much space. Therefore, we create a special box where we explain the methodology. And in the story, we refer to the box. (CH)

> For almost every story we have a "read me" file that kind of explains, this is what we reported, this is what data we used. And sometimes it'll have extra contextual stuff where we've had those sorts of discussions; there will be an explanation of the methodology if you like. But that isn't in the body of the article (…). (UK)

Rinsdorf and Boers (2016) regard transparency – in addition to fact checking – as a strategy of quality management: allowing readers to sift through the data, check what journalists have done, evaluate the methods applied, and verify their findings. Even readers who will not spend time on such activities may appreciate being positioned as qualified for doing so (cf. Engebretsen, 2017).

> Sometimes we publish the complete source code, the analysis script to show how we processed the data so that the user can understand our approach and maybe detect errors. (CH)

> What we are doing in this project for the first time is that we make the R-code available with R Markdown, in addition to the raw data. And we explain why we do this. (DE)

> Another thing is to be open in a manner that makes readers actually eager to check out what we have done. One possible way is, I guess, to state clearly how we have used the data, and what we have chosen as an angle to the story. And we could also approach this similar to what is done in research, constantly ask ourselves, "Is this correct?," and then show alternative ways to interpret and visualize the numbers. (NO)

Another aspect of the meta-story is how data stories are shared, re-shaped and continued in discussions and commentaries on social media. Most of the interviewees agreed that social media are an essential part of their work. They produce special graphics or animated gifs for social media to draw people into the data story, to increase click rates and to trigger discussions.

> We always distribute a gif or a screenshot on Twitter or Facebook to promote our interactive visualizations so that people realize: hey, here is something visual to discover. (CH)

> It should have a wow effect! Something that has come lately is the explainer in video format, presented on Facebook or Youtube. Here, the graphics are embedded in the video and it is all perceived as a linear narrative. It happens more and more. An important reason for this is that video is so easy to share on social media. (NO)

However, the benefits or effects of sharing graphics on social media are not part of this study.

5 Conclusion

New technologies and the opening up of public data have brought data into newsrooms, inviting journalists to tell stories with data and through data visualizations. The current prevalence of data in Western culture is pushing numbers as a semiotic resource – besides text and images – towards the center of journalistic practices. In data stories, data plays a core role, and the patterns, outliers, relationships, correlations, or differences found in the data act as a stimulus to initiate the process of storytelling. The interviewees called this kind of stories "data-driven, data-informed stories." To see and show what is striking in the dataset, journalists need to visualize the data. Kirk (2016, p. 21) calls this process of making data visible "the act of representation." This process of visualization can lead to a simple bar chart, a map, a stand-alone visualization, or a combination of text and visualizations.

Data journalists aim to convey a clear message through the act of showing through visualization. This act of showing consists of narrative, explanatory, and argumentative elements and structures. In addition, data stories often enable users to choose their own story, their individual angle on the story by exploring the data by themselves. As the interviewees pointed out, one single data visualization "can tell different stories, not only one." Looking for the story most personally relevant triggers involvement. Offering multiple micro stories, the Martini glass structure with its linear-nonlinear dramaturgy, exemplifies this, a compromise between an author-driven and a reader-driven structure: Firstly, data journalists handle datasets more from distance, which means from an authorial point of view offering one or *the* basic story; the aim is to show the users how to read the visualization and not to overwhelm them with data. Secondly, the data visualization opens up many possible individual stories with a more personal narrative point of view. Providing data through data visualizations calls for transparency regarding journalistic practices and editorial processes. Transparency as a qualitative management strategy and an ethical standard in data journalism means explaining the process of data collection, analysis, and presentation and allowing users to check the data work of the journalists. Last but not least, another aspect of visual communication is gaining center stage in journalistic storytelling: design – how the data is presented. To attract readers' interest on a website or on social media, data stories with a compelling visual design are more successful than text-centric stories. By tweeting or posting a screenshot of a data visualization on social media, the visibility of a data story, and thus, of a news organization, can be increased.

Although data visualizations often constitute or are included in textual structures with elements of narrativity, there are reasons to discuss whether narrative is the most relevant term for the conceptualization of exploratory data visualizations. A basic characteristic of narrative in a traditional sense, is, as earlier noted, the unbroken chain of events, where temporal and causal relations create the coherence of the story (Bell, 1991). This is hardly the case when the communication is user-driven rather than author-driven. Further, the term narrative will, in traditional text theory, often be related to a distinct voice, a narrator. Genette (1980) refers to the platonic distinction between diegesis and mimesis when he talks about "pure narrative" and "perfect imitation" in literary works. This distinction is closely related to what we in everyday speech refer to as "telling and showing," and which in the context of journalism education is often found in the form of normative advice: "Show, don't tell!" (e. g., Mencher, 1997, p. 154). Following this line of thought, we can suggest that data-driven news stories structured around the Martini glass model result in a mix of telling and showing, diegesis and mimesis, although the mimesis

part – the imitation of the world – is of a highly abstracted kind.

To sum up, we identified seven key features of journalistic data stories. These are: data as the core of a story, the diverse communicative purposes, the new textual-visual relationships, new structures and forms of story design, different kinds of interactivity, and the increasing importance of the meta-story. Given these key features, it is definitely the case that understandings of journalistic storytelling need to be re-thought. We might not want to go as far as saying that a narrative approach to journalistic practice is not at all useful. But our interview analysis has shown that besides narration there are other textual practices, visual elements, structural patterns, and interactive techniques in journalistic artifacts today which require us to rethink what we understand as journalistic storytelling such that we have to either (i) update our understanding of what counts as a story or (ii) stop calling journalism stories. Either way, it is clear that traditional definitions of stories do not apply in the contemporary news production landscape.

Our empirical findings can be seen as a first step to rethinking the narrative approach. It can also be seen as a contribution to understanding a bigger picture of the significant changes that journalistic practices and newsmaking are undergoing. To enlighten the discussion on whether data stories fulfill the criteria for a story as discussed in narrative theory, and in what ways they represent new narrative forms and functions, we need further analysis of the artifacts themselves, the data stories. This is the next step for research in this area.

Acknowledgements

The study is part of the research project, Innovative Data Visualization and Visual-Numeric Literacy (INDVIL, www.indvil.org). The research project INDVIL is supported by the Research Council of Norway (NFR) and the Norwegian Media Authorities (RAM).

References

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2017). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*. Advance online publication, 1–24. doi:10.1177/1464884917700667.

Bell, A. (1991). *The Language of News Media.* Oxford, UK: Blackwell.

Bertini, E., & Stefaner, M. (n. d.). *Data Stories. A podcast on data visualization with Enrico Bertini and Moritz Stefaner.* Retrieved from http://datastori.es/.

Borgo, R., Kehrer, J., Chung, D. H. S., Maguire, E., Laramee, Robert S., Hauser, H., Ward, M., & Chen, M. (2013). Glyph-based visualization: Foundations, design guidelines, techniques and applications. *Eurographics State of the Art Reports,* 39–63. doi:10.2312/conf/EG2013/stars/039-063.

Bradshaw, P. (September 14, 2017). *No, I'm not abandoning the term "storytelling," Alberto – just the opposite (and here's why).* Retrieved from https://onlinejournalismblog.com/2017/09/14/narrative-storytelling-data-journalism-alberto-cairo/.

Brehmer, M., Lee, B., Bach, B., Riche N. H., & Munzner, T. (2016). Timelines revisited: A design space and considerations for expressive storytelling. *IEEE Transactions on Visualization and Computer Graphics, 23*(9), 2151–2164. doi:10.1109/TVCG.2016.2614803.

Brinker, K. (2010). *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden.* [Linguistic text analysis. An introduction … (7th vol.) Berlin: Erich Schmid.

Cairo, A. (2016). *The truthful art. Data, charts, and maps for Communication.* Berkeley, CA: New Riders

Cairo, A. (2013). *The Functional art: An introduction to information graphics and visualization.* Berkeley, CA: New Riders.

Card, S. G., Mackinlay, J., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think.* San Francisco, CA: Morgan Kaufmann.

Engebretsen, M. (2017). Levende diagrammer og zoombare kart. Datavisualisering som nyskapende fortellerform i journalistikken

[Live charts and zoomable maps. Data visualization as an innovative narrative form in journalism]. *Norsk Medietidsskrift, 2*(24). doi:10.18261/issn.0805-9535-2017-02-02.

Engebretsen, M. (2006). Shallow and static or deep and dynamic? Studying the state of online journalism in Scandinavia. *Nordicom Review, 27*(1), 3–16.

Engebretsen, M., & Weber, W. (2017). Graphic modes. In C. Cotter, & D. Perrin (Eds.), *Handbook of language and media* (pp. 277–295). London: Routledge.

Engebretsen, M., Kennedy, H., & Weber, W. (2017). Visualization practices in Scandinavian newsrooms: a qualitative study. In E. Banissi et al. (Eds.), *21st International Conference Information Visualisation (IV 17). Computer graphics, imaging and visualisation* (pp. 297–300). London: IEEE Computer Society. doi: 10.1109/iV.2017.54.

Forster, E. M. (1974). *Aspects of the novel*. London: Arnold. [1927].

Genette, G. (1980). *Narrative discourse*. Oxford: Blackwell.

González Veira, X. (September 13, 2017). *I like data-driven storytelling*. Retrieved from https://medium.com/xocas/i-like-data-driven-storytelling-i-like-loaded-words-d066d83ee909.

Hamilton, J. (2016). *Democracy's detectives: The economics of investigative journalism*. Cambridge, MA: Harvard University Press.

Hermida, A., & Young, M. L. (2017). Finding the Data Unicorn. A hierarchy of hybridity in data and computational journalism. *Journalism, 5*(2), 159–176, doi:10.1080/21670811.2016.1162663.

Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2231–2240.

Hullman, M., Kosara, R., & Lam, H. (2017). Finding a clear path: Structuring strategies for visualization sequences. *Computer Graphics forum, 36*(3), 365–375. doi:10.1111/cgf.13194.

Jahn, M. (2017). *Narratology: A guide to the theory of narrative*. English Department, University of Cologne. Retrieved from http://www.uni-koeln.de/~ame02/pppn.htm.

Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication and Society 19*(6), 715–735.

Kirk, A. (2016). *Data visualisation: A handbook for data driven design*. Los Angeles, London: Sage.

Kosara, R. (2017). An argument structure for data stories. Short Paper. *Proceedings of the Eurographics.IEEE VGTC Symposium on Visualization (EuroVis)*.

Kress G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. New York: Routledge.

Marconi, F., Siegman, A., & Machine Journalist (2017). The future of augmented journalism. A guide for newsrooms in the age of smart machines. *AP Report*. Retrieved from https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf

Matzat, L. (2016). *Datenjournalismus*. Konstanz: UVK.

Mayring, P. (2000). Qualitative Content Analysis [28 paragraphs]. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research, 1*(2), Art. 20. Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204.

Mencher, M. (1997). *News reporting and writing* (Seventh Edition). London, Chicago: Brown & Benchmark.

Merminod, G. (2016). Telling stories from the newsroom: A linguistic ethnographic account of dramatization in the broadcast news. *Working Papers in Urban Language & Literacies*. Paper 197. Retrieved from http://www.academia.edu/28219120/WP197_Merminod_2016._Telling_stories_from_the_newsroom_A_linguistic_ethnographic_account_of_dramatisation_in_the_broadcast_news.

Perrin, D. (2015). Die Kunst des In-die-Geschichten-Erzählens: Der Reiz der Unvollendeten. In M. Prinzing (Ed.), *Die Kunst der Story* (pp. 95–111). Thun/Gwatt: Werd & Weber Verlag AG.

Perrin, D. (2011). "There are two different stories to tell": Collaborative text-picture production strategies of TV journalists. *Journal of Pragmatics, 43,* 7. 1865–1875.

Rinsdorf, L., & Boers, R. (2016). The need to reflect. Data journalism as an aspect of

disrupted practice in digital journalism and in journalism education. *Proceedings of the International Association for Statistical Education*. IASE Roundtable 2016 Berlin: Promoting Understanding of Statistics about Society. Retrieved from http://iase-web.org/documents/papers/rt2016/Rinsdorf.pdf.

Rogers, S., Schwabish, J., & Bowers, D. (2017). *Data journalism in 2017. The current state and challenges facing the field today.* Google News Lab, September 2017. Retrieved from https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf.

Ryan, M.-L. (2006). *Avatars of story.* Minneapolis, MN: University of Minnesota Press.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics,* Volume 16, Issue 6, November 2010, 1139–1148. DOI 10.1109/TVCG.2010.179.

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2016). *Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories.* Technical report, Microsoft Research. Retrieved from http://research.microsoft.com/apps/pubs/default.aspx?id=264484.

Sunne, S. (September 3, 2016). *How data journalism is different from what we've always done.* Retrieved from https://www.americanpressinstitute.org/publications/reports/strategy-studies/how-data-journalism-is-different/.

Van Leeuwen, T. (2005). *Introducing social semiotics.* London: Routledge.

Weber, W. (2017). Interactive information graphics: A framework for classifying a visual genre. In A. Black, P. Luna, O. Lund, & S. Walker (Eds.), *Information design: research and practice* (pp. 243–256). Abingdon, New York: Routledge.

Weber, W. (2013). Typen, Muster und hybride Formen. Ein Typologisierungsmodell für interaktive Infografiken. In W. Weber, M. Burmester, & R. Tille (Eds.), *Interaktive Infografiken* (pp. 25–37), Heidelberg, Berlin: Springer.

Weber, W., & Rall, H.-M. (2016). Design and journalism. Challenges and opportunities: A dialogue between two cultures. In W.W. Wong, Y. Kikuchi, & T.S. Lin (Eds.), *Making Trans / National Contemporary Design History. 10th Conference of the International Committee for Design History & Design Studies* (ICDHS 2016) (pp. 313–318). Sao Paulo: Blucher.

Weber, W., & Rall, H.-M. (2012). Data visualization in online journalism and its implications for the production process. In E. Banissi et al. (Eds.,) *Information Visualisation. Proceedings of the 16th International Conference on Information Visualisation* (pp. 349–356). Montpellier, France: IEEE Computer Society. DOI 10.1109/IV.2012.65.