Developing Data Stories as Enhanced Publications in Digital Humanities

Willemien Sanders, Roeland Ordelman, Mari Wigham, Rana Klein, Jasmijn Van Gorp, Julia Noordegraaf

Work in progress - please do not cite without permission from the authors

Abstract

This paper discusses the development of data-driven stories and the editorial processes underlying their production. Such 'data stories' have proliferated in journalism but are also increasingly developed within academia. Although 'data stories' lack a clear definition, there are similarities between the processes that underlie journalistic and academic data stories. However, there are also differences, specifically when it comes to epistemological claims. In this paper data stories as phenomenon and their use in journalism and in the Humanities form the context for the editorial protocol developed for CLARIAH Media Suite Data Stories.

Key words

data stories, storytelling, digital journalism, digital humanities, data visualizations, data narratives

Introduction

The datafication of society and – more specifically – of cultural heritage collections has continued to thrive over the past two decades. Collection owners have developed infrastructures to make digitized or born digital collections accessible and to allow activities such as the bookmarking, annotating and analysing of collection content. The many publications discussing Digital Humanities testify to this thriving field of inquiry. One example of such an infrastructure is the CLARIAH Media Suite, developed to provide access to multimedia collections and hosted by the Netherlands Institute for Sound and Vision. The Media Suite currently offers 91 datasets from 62 cultural heritage and knowledge organizations in the form of both data and metadata. This includes, among others, large bodies of historical radio, television, film, and newspaper data. Within CLARIAH, the Common Lab Infrastructure for Arts and Humanities, the Media Suite has been developed for and with media and humanities scholars to facilitate searching and working with the collections it holds (see, for instance, Melgar et al. 2017, Melgar-Estrada et al. 2019, Ordelman et al. 2018, Van Gorp, Melgar Estrada, and Noordegraaf 2021).

While the development of the Media Suite is founded on the needs of media scholars trained primarily in qualitative methods, the large dataset that the Media Suite

provides access to are also regarded as a valuable potential source for quantitative research for both media scholars and humanities scholars in general. As Windhager et al. (2019) argue, the growing collections invite new ways for analyzing and understanding those collections. That is why we have started experimenting with data-driven research and publications: the Media Suite Data Stories. These Data Stories can be understood as an additional layer on top of the Media Suite, a layer which facilitates the publication of – mainly – quantitative research with the Media Suite data and metadata. In addition, and making this a rather unique form of publication, the underlying data are accessible, providing an opportunity for readers to scrutinize the authors' work. So far, Media Suite Data Stories have been produced by both scholars and professionals, including cultural heritage institute professionals and journalists. This paper focuses on academic Media Suite Data Stories.

There is a popular idea that "some of today's most relevant stories are buried in data" (Riche et al. 2018;4, see also Figueiras and Vizoso 2022). Considering the problem with data bias and the challenges of understanding data properly (of which power imbalance between those usually collecting data and those usually represented in the data is not the least one, see D'Ignazio and Klein 2020), ideas like these need critical scrutiny. Although D'Ignazio and Klein (2020) refer to data analysis as 'story finding', which likewise reflects the idea that stories are to be found in data, this also relates to the practice of exploring data. We will come back to this below.

Fickers and Clavert (2021:par 10) argue that "we are in desperate need for clearer protocols on how to document and share self-reflexive position statements in scientific publications online", in other words, how to account for our work in, for instance, publications like Media Suite Data Stories. In addition, D'Ignazio and Klein (2020) point to the need for contextualization: information about how and why data are collected and how they are processed - in other words, how data sets are "produced" (Kitchin 2021: 5) - before they are turned into a dataset ready for use. According to Condie and Costa (2018:206) it is crucial to address how projects "recognize the interplay between technology and research" and to interrogate research practices and assumptions. They discuss the three defining aspects relevant to any type of research: the researcher's ontological, epistemological, and methodological stance. With this paper we hope to contribute to this discussion. In this paper we argue there are specific demands for the production process of a data story as an academic argument, and discuss the editorial protocol we developed for making Media Suite Data Stories.

In what follows we will discuss the onset of the phenomenon of data-driven stories in the wider context of both journalism and academia, different initiatives within the Digital Humanities, the challenges of working with the Media Suite data, and the

editorial protocol we developed to safeguard the quality of both the content and technology of Media Suite Data Stories. We conclude with ideas for future work.

Data stories and their (lack of) definition

According to Riche et al. (2018) the development of data-driven stories occurred in both academia and journalism, although the former tends to interrogate the possibilities and potential of visualizations while the latter are "at the forefront" (5) of the production of data-driven stories. The term 'data stories' has been used in a variety of publications to refer to different narratives involving quantitative data and, usually, visualizations, especially in journalism (Weber 2020). However, there is a lack of definition of what makes stories or data driven output a 'data story', and a lack of discussion of the relationship between data visualisations and storytelling.

Some researchers discuss data stories in terms of their production. Gray and Bounegru (2021:3) provide an explanation in parenthesis, and limit it to "collecting, analyzing, visualizing and narrating data", thereby circumventing a discussion. Arguably, apart from visualizing, this reflects the process of any research project, qualitative or quantitative in nature. Weber, Engebretsen, and Kennedy (2018) discuss two definitions of data journalism. The first relies on work by Howard (2014: 4): "gathering, cleaning, organizing, analyzing, visualizing and publishing data to support the creation or acts of journalism". The second relies on work by Rinsdorf and Boers (2016: 1): "analyzing open data sets using (semi-)automatized methods to detect meaningful patterns in data structure". Both focus on data and their analysis; neither explicitly refers to telling stories. Weber (2020:296) also refers to 'data storytelling' as a "structured approach" for the communication of insights from data and as involving data, visuals and narrative. This approach suggests a systematic workflow.

As D'Ignazio and Klein (2020) argue, numbers do not speak for themselves and context is key to understanding them. In order to be able to tell stories based on data, it is necessary to be or become acquainted with those data and understand what they represent. That is why an exploration of the data should be part of any research effort too, an exploration that goes beyond 'cleaning'. Such an exploration should be geared towards understanding what the data are about, what the possibilities and limitations are, and which analyses seem fruitful.

Weber, Engebretsen, and Kennedy (2018: 194) argue that the term 'data story' suggests a narrative, which, relying on work by Bell (1991), they define as "a sequence of events temporally structured and coherently related to each other with bonds of (strong or weak) causality" (see also Figueiras and Vizoso 2022, Weber 2020). This reflects a common understanding of 'narrative', in which events, time and causality are central elements. According to Weber, Engebretsen, and Kennedy (2018), journalistic

storytelling involves narrating, describing, explaining and arguing. The same could be said about academic stories or arguments, specifically in the Humanities.

According to Riche et al. (2018), data-driven stories "start from a narrative that either is based on or contains data and incorporates this data evidence … to confirm or augment a given story" (8), often in the form of visualizations. Data visualizations help to take the reader through the information and knowledge gained from the data and can provide context, they argue. Accordingly, Weber, Engebretsen, and Kennedy (2018:194) argue that data visualizations "amplify the cognitive processing and the social application of the data represented". Duangphummet, Apiwan, and Puripant Ruchikachorn (2021:242) state that "storytelling allows data visualization to reveal analysis results compellingly and effectively". In other words, visualizations help to understand the research results, arguably for journalism and academia alike. Figueiras and Vizoso (2022: 87) argue that although visualizations usually support "traditional storytelling forms", they increasingly form independent stories. However, very few authors explicitly develop ideas about story and storytelling with such visualisations. They usually discuss the use of various techniques to facilitate storytelling.

Others discuss data stories in terms of the nature of the output of data-driven research. Based on the work of a number of other scholars, Weber, Engebretsen, and Kennedy (2018:192) describe data stories as multimodal and hybrid artefacts that can be based on numbers, texts and images to create "a coherent whole". Here the focus is on the combination of various textual forms and communication strategies. According to Weber, Engebretsen, and Kennedy (2018:192), data visualizations are central to data stories, visualizations "that range from simple to complex multimodal interactive stand alone graphics". Later on, they refer to data stories as "storytelling with data" (193). As argumentation is key to the Humanities, it is relevant to discuss the relationship between visualization and narrative.

Based on the above we define 'data stories' as follows: Data stories are the output of an iterative process of data collection, data exploration, data preparation, data analysis, data visualization, interpretation and narration. The result is a narrative that is based on the analysis of quantitative data and includes visualisations of these data with explanations of their meaning and their context.

As mentioned above, data stories have predominantly been developed within journalism (Riche et al. 2018). As a consequence, research on storytelling with data has proliferated in journalism studies. That is why we turn to these debates first, to understand the similarities and argue the differences between journalistic and academic data stories.

Data stories in journalism

According to Arrese (2022) the use of data for journalism can be traced back to early economic journalism in the 17th and 18th centuries. This developed into data-driven reporting in the 19th century. With the growing belief in numbers and measurements as well as the development of digital infrastructures and services, data have become increasingly popular in other fields as well and so journalists widened their scope to report data-based news in a wide variety of fields. Data stories have become widely accepted in journalism as the output of a process that is in many aspects similar to that of a scholar in terms of their production.

Journalists partly rely on their own data collection, for instance, of social media data, but they also partly rely on data provided by organisations, including government agencies. Those data, before being offered to outsiders to work with, have been collected and processed. This includes, amongst others, parsing, filtering, and refining (Fry 2008). Outliers and missing data might also have been dealt with, amongst others. As mentioned above, for anyone working with large datasets, it is important to consider the context of those data – how and why they were collected and how they have been processed. Such knowledge is vital for the epistemological claims journalists can make.

According to D'Ignazio and Klein (2020), such public data is often offered without context, or with only minimal context, as governments support the sharing of data but allocate insufficient resources to provide proper context about the collection and processing (or cleaning up) of the data. Therefore, it is hard to properly understand such data and interpret the results of analyses. Such data acquire a "quiet authority" (Lowrey, Broussard, and Sherrill 2019:70), and As we will see, knowledge of the context of data, their provenance, the construction of the dataset and the limitations of it, are key to understanding any analysis conducted with it.

Weber, Engebretsen, and Kennedy (2018) argue that data visualization calls for transparency regarding practices and editorial processes, both as qualitative management strategy and as ethical standard. This includes explaining the collection, analysis, and presentation of data and allowing users to check these practices. This aligns with academic practices of accountability. However, with data sets provided by others, explaining the selection might be hard, if not impossible.

Most work on data visualizations in journalism discusses visualizations as a means to tell stories. By including a narrator, sequentiality, time and "tellability" (Weber 2020:307), data visualizations become narratives. To understand how journalists make use of data, scholars have analysed journalistic 'data stories'. For instance, Segel and Heer (2010) have analysed 58 data based stories and visualisations and provided a taxonomy including seven genres which can be positioned on an axis between author-driven and reader-driven narratives. The former are more explanatory in nature, the

latter more exploratory (see also Weber 2020). These seven genres are: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation. Further developing this work, Stolper et al. (2016) distinguish between four broad non-exclusive categories: communicating narrative and explaining data; linking separated story elements; enhancing structure and navigation; and providing controlled exploration. Evidently, this includes a wide variety of ways on which data can be visualized and stories can be told.

As a result, we might distinguish between this practice of storytelling *through* visualizations and storytelling *with* visualization, which seems more suitable to current practices in the Humanities. In the latter, the visualizations are part of a larger story in writing that explains and contextualizes the research the story is reporting.

Data stories in the Humanities

In the Digital Humanities there are a number of initiatives that aim to publish (data-driven) research in a format that addresses the demands of Humanities scholarship, sometimes referred to as enhanced publications: publications enriched with or linked to related research results (Bardi and Manghi 2015). For instance, Scalar (scalar.me) is a platform for enriched multimedia publications, facilitating multi-layered publications that might include both the central argument as well as link to the underlying material it is based on, through the import of metadata. The Journal of Digital History (journalofdigitalhistory.org) is characterized by "multilayered publication" for data-driven research and "transmedia storytelling", as different layers with different media can be included. Related initiatives include the Journal of Open Humanities Data (openhumanitiesdata.metajnl.com), which focuses on the sharing and reuse of data and techniques, and the Journal of Data Mining & Digital Humanities (jdmdh.episciences.org), which positions itself at the intersection of computing and the Humanities.

The Journal of Digital History publishes data-driven research through the "new digital practice of writing, visualizing, and arguing history" (Fickers and Clavert 2021:par 16). According to Fickers and Clavert (2021), in their editorial to the first issue of the journal, this practice might include "producing transparency about how the digital interferes in the iterative process or lifecycle of the research process" which may be considered an "epistemological imperative". Bardi and Manghi (2015) argue that, as publications are only released at the very end of the research process, data and their processing methods should be shared as well in order to meet academic standards of replicability. Likewise, Fickers and Clavert argue the potential of opening up the research process to readers by providing not only the results of a given research project but also the process itself. The “weaving [of] interpretation, narrative, evidence, and

commentary" (par.4) allows the reader to "think along" with the goal to be transparent about the way digital technologies interact with the research process, as this has epistemological consequences.

Indeed, for Humanities scholars it is adamant that they are clear about their research practices and activities. While journalists use data research mainly to uncover more or less hidden truths believed to be unaffected by journalists' values (Parasie 2015), and thereby approach the results of the analysis of data as reflecting a truth that speaks for itself, Humanities scholars are used to constructing an argument, a convincing narrative from a position of situated knowledge (Haraway 1988), and the results of data analyses are extensively contextualized to make the argument convincing, reliable and reproducible. In other words: the digital scholarly practice might include transparency with respect to how scholars include the digital data and tools they use into their research practices (see Condie and Costa 2018), because such new digital practices have consequences for the epistemological claims scholars may present and that results from their situated position.

Figueiras and Vizoso (2022) discuss the process of creating 'visual data stories' as a decision making process with several iterations, for which guidelines might be helpful. Duangphummet, Apiwan, and Puripant Ruchikachorn (2021) provide valuable insight into their experiences in creating data visualizations and stories and the advancing insight they developed. They describe what they learned in different phases of developing various data visualizations, which comes close to describing their editorial insights and decisions. Hence, Duangphummet, Apiwan, and Puripant Ruchikachorn share their lessons learned from their interdisciplinary collaboration. While working on a series of visualizations they updated the interdisciplinary team. It was only after the first iteration that they included a domain expert in the team. Later they also involved a data scientist and analyst. The protocol Duangphummet, Apiwan, and Puripant Ruchikachorn developed includes the following stages: conceptualization, data preparation, realization, visualization design, and visualization development. Their work focuses on visualizations but mirrors the production of data-based narratives.

The above sketches the context within which we are developing a publication infrastructure on top of the Media Suite. Below we will discuss the Media Suite Data Stories, followed by a discussion of the editorial protocol we developed for their creation.

Media Suite Data Stories

In addition to providing the Media Suite as a tool for browsing, searching and analysing, we have been collaborating with both humanities scholars and journalists to develop "Media Suite Data Stories". These data stories align with what Stolper et al. [8] refer to as "author-defined" (1) narratives, and are further characterized by (a) the combination

of domain specific knowledge and a critical approach to data and tools, (b) transparency with respect to data analysis, and (c) accessibility through multi-layering of the publications: through a Data Story, the underlying data – i.e., the contents of the Media Suite, which is subject to copyright – becomes accessible.

Media Suite Data Stories combine visualizations of data analyses with texts that explain those visualizations and connect them into a narrative. Media Suite Data Stories are published as web pages and the reader can scroll down as s/he progresses through the story. Publishing stories this way is also referred to as 'scrollytelling' (Weber 2020). Stolper et al. argue that scrolling is "a pervasive and powerful technique used in data-driven storytelling" (12 / 6.3) even if a better understanding of its strengths and weaknesses is needed.

Media Suite Data Stories are based on data and metadata from the Media Suite. The datasets available in the Media Suite come with extensive explanation, offered through the Media Suite data registry (https://mediasuitedata.clariah.nl). Here, for each collection there is information about its size, the kinds of media it includes and other characteristics. This gives users a minimum of necessary information about the data available and thus contextualizes the data (D'Ignazio and Klein 2020). Unfortunately, as data processing from, for instance, the national public broadcaster NPO to the Media Suite remains a black box for many, creating full transparency about the data is not a matter of course. For instance, many metadata fields are completed by NPO staff and their reasoning for specific metadata entry is not always clear. In addition, the complex processes of data transfer from NPO to Sound and Vision sometimes gets interrupted, which threatens the availability of complete datasets. At the same time, in the Media Suite Data Stories, the production (e.g., through ASR, voice recognition, and face recognition), selection, processing, analysis and visualisation of data are all accounted for and available for scrutiny.

However, according to Leon (2021), there are limits to what authors (be it journalists or academics) can explain in writing, specifically when it comes to explaining data analysis. He proposes to publish code, specifically in literate programming languages such as Jupyter Notebooks (also used by the Journal of Digital History) to not only make the process and steps in the analysis explicit, but also to streamline analysis and quality control procedures, especially when the source data are unavailable for sharing. Media Suite Data Stories are also produced with the help of Jupyter Notebooks. Unfortunately, due to copyright restrictions on the data in de Media Suite, these are not yet widely available.

According to Fickers and Clavert (2021), workflows describe the experimental process of knowledge production in a formalized way (par. 10). Based on our experiences we developed an editorial workflow or protocol which includes four

(iterative) stages and is aimed at safeguarding the quality of both the domain specific knowledge transfer and a critical approach to digital data and tools, as well the transparency of the research process and outcomes. By providing insight into the editorial process, we not only hope to develop the discussion by Fickers and Clavert as well as Duangphummet and Ruchikachorn (2021) concerning digital research practices referred to above, but also take a step towards lowering the threshold for prospective researchers interested in conducting and publishing research based on multimodal media data and metadata.

The editorial process for making Media Suite Data Stories is facilitated by a core team and consists of four phases: exploration, research, review, and publication. We will discuss each in turn below.

Exploration

Media Suite Data Stories are based on - and their production driven by - a research question or hypothesis. This is because the available data in the Media Suite is too extensive and diverse to just 'see what you find'. In addition, an open exploration to find a story runs the risk of finding different results and connecting them without justification. Therefore, we use exploration to develop a research question or hypothesis and test the dataset, of which understanding it (D'Ignazio and Klein 2020) is a part. Based on a data collection and a research interest the data are explored in order to come to a research question or hypothesis and a suitable delineation. Four steps are undertaken to get there. First, the researcher and data scientist brainstorm about the research interest and potentially useful data. A data scientist and a data engineer are part of the core editorial team of Media Suite Data Stories to guarantee that knowledge about the (meta)data is present in each project. Second, they translate their ideas into measurable questions, suitable for quantitative research. Third, they execute a preliminary research project, like a pilot study, to test their ideas and data set. And fourth, they discuss the results to see if they can proceed to the next phase. Like all phases in the Media Suite Data Story process, this one is iterative, and the team might need several iterations to come to a feasible research plan. Once they feel the research question/hypothesis and data set match sufficiently and provide promising first results, they move to the next phase.

Research

The research phase is aimed at investigating the research question/hypothesis through a Data Story. The results form the basis for the story but the progress of the research depends to some extent on the results of consecutive analyses. That is, based on the first analyses and visualizations thereof, relevant patterns or developments may occur

that deserve further investigation. As a result, in this phase the narrative of the story is also constructed. To get there, the team takes five steps in this phase. First, they create focus in their project by reconsidering and selecting specific questions/hypotheses to focus on and/or reformulating them to make them more specific. Second, they operationalize these questions/hypotheses so that they can be investigated. Third, they generate the necessary data and analyse them (in line with Duangphummet and Ruchikachorn's (2021) data preparation). Fourth, they interpret the results, and fifth, they tell the story. Again, these steps are iterative and going back and forth will be necessary to properly develop a sound Media Suite Data Story. Once the analyses have been completed and the story written, it is time to move on to the next phase.

Review

The goal of the review phase is to ensure both the technological quality of the Media Suite Data Story and the quality of its narrative. Fickers and Clavert (2021) point to the challenge of peer review and refer to it as the evaluation of "the results of a research, its methodology, its code and its data". Unfortunately, Fickers and Clavert do not elaborate on how they went about this and for now it remains "not fully solved" (par 29). For Media Suite Data Stories, we invite experts not involved in the production and/or writing of the data story to review them. Domain experts review the story's content: its narrative and the conclusions based on the analyses. They then deliver recommendations with respect to the research domain. Technical reviewers focus on the data used for the research, the code, the analyses, the connection with the underlying data, and deliver recommendations with respect to the technical aspects of the Media Suite Data Story. In addition, we consider whether there are any legal aspects that need attention (such as potential copyright infringements or privacy issues). The resulting recommendations are considered by the team and implemented where deemed desirable and feasible. Once this process is finished, the final phase starts.

Publication

The publication phase is aimed not only at publication itself but also at checks, promotion and monitoring of the Media Suite Data Story. To this end, the Media Suite Data Story is first published on a dedicated platform (https://mediasuitedatastories.clariah.nl), after which the layout is checked for errors. Once approved, efforts are made to publicize the Media Suite Data Story through means and channels relevant to the research domains and related communities. This may include press releases, newsletters, social media posts and other means. Finally, we monitor responses to the Media Suite Data Stories.

Conclusion and future work

At the time of writing, five Media Suite Data Stories have been published and we are working on another three. We use the protocol described above as a framework for the work and we intend to update it where necessary. As Fickers and Clavert (2021) have acknowledged, the creation of these kinds of research narratives is a lot of work (see also Figueiras and Vizoso 2022 with respect to journalistic data visualizations). We are continually learning and evaluating to find the best way forward.

One of the main challenges we face is making the creation of Media Suite Data Stories more appealing to academics. Researchers in the Humanities are still mainly trained in qualitative research methods and taking a quantitative approach is not a matter of course, as we experienced. Our main aims for the future involve lowering the threshold for scholars to create Media Suite Data Stories. At the same time, we aim to develop more visual styles of communication.

To lower the threshold for academics we are working on three ideas. First, we want to develop modules or building blocks for Media Suite Data Stories. By modules or building blocks we mean specific queries, analyses and/or visualizations that can be adapted to researchers' needs without much knowledge of code. For instance, for a Media Suite Data Story on the discourse on 'fake news' in the Netherlands we created a query that specifically searches for the eight o'clock news broadcasts of the Dutch public broadcaster, a challenge given the many ways in which these programmes have been archived in terms of metadata. Considering that future researchers might well be interested in including this selection in their research, we saved this query for future use. Researchers may then reuse it with the option to change, for example, the period they wish to research. Another example is the analysis of the occurrence of a specific term, or a combination of terms, in this collection of newscasts resulting from the query. Researchers may, without much hassle, adapt the terms they want to count rather than write new code. To offer such building blocks, for which we find Jupyter Notebooks useful (see Leon 2021), we are working on creating an environment in which researchers may use Jupyter Notebooks on Media Suite data, which are copyright protected. Such building blocks also help to create a shared epistemological ground, as researchers reuse existing methods in the form of data collection and analysis.

Second, we aim to recruit scholars experienced in data research and/or teams of researchers and data scientists. By recruiting researchers or research teams more experienced in quantitative and data research we wish to make the production less resource intensive for the core team and facilitate the creation of more stories. And third, we wish to develop a pilot project in collaboration with an academic journal, to 'proof the concept' of academic Media Suite Data Stories. By exploring the possibility for a Media Suite Data Story as a proper academic article developed in collaboration with an

academic journal, we hope to help develop multimedia academic storytelling and facilitate the creation of future academic Media Suite Data Stories.

Acknowledgement

This work was made possible by the CLARIAHPLUS project funded by NWO (Grant 184.034.023).

References

Arrese, Ángel. 2022. '"In the Beginning Were the Data": Economic Journalism as/and Data Journalism'. *Journalism Studies* 23 (4): 487–505.

Bardi, Alessia, and Paolo Manghi. 2015. 'A Framework Supporting the Shift from Traditional Digital Publications to Enhanced Publications'. *D-Lib Magazine* 21 (1/2). Available at https://www.dlib.org/dlib/january15/bardi/01bardi.html.

Bell, Allan. 1991. *The Language of News Media*. Language in Society 16. Oxford, UK ; Cambridge, MA: Blackwell.

Condie, Jenna, and Cristina Costa. 2018. '(Re-)Exploring the Practical and Ethical Contexts of Digital Research'. In Doing Research In and On the Digital: Research Methods across Fields of Enquiry, edited by Cristina Costa and Jenna Condie, 1st ed., 205–12. London: Routledge.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Strong Ideas Series. Cambridge, MA: The MIT Press.

Duangphummet, Apiwan, and Puripant Ruchikachorn. 2021. 'Visual Data Story Protocol: Internal Communications from Domain Expertise to Narrative Visualization Implementation': In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - IVAPP*, 240–47. SciTePess.

Fickers, Andreas, and Frédéric Clavert. 2021. 'On Pyramids, Prisms, and Scalable Reading'. *Journal of Digital History*, no. jdh001 (October). Available at https://journalofdigitalhistory.org/en/article/jXupS3QAeNgb.

Figueiras, Ana, and Ángel Vizoso. 2022. 'Infomation Visualization: Features an Challenges in the Production of Data Stories'. In *Total Journalism: Models, Techniques and Challenges*, edited by Jorge Vázquez-Herrero, Alba Silva-Rodríguez, María-Cruz Negreira-Rey, Carlos Toural-Bran, and Xosé López García, 83–96. Cham, Switzerland: Springer.

Fry, Ben. 2008. *Visualizing Data. Exploring and Explaining Data with the Processing Environment*. Sebastopol, CA: O'Reilly Media, Inc.

Gray, Jonathan, and Liliana Bounegru. 2021. 'Introduction'. In *The Data Journalism Handbook: Towards A Critical Data Practice*, edited by Liliana

Bounegru and Jonathan Gray, 11–23. Amsterdam, The Netherlands: Amsterdam University Press.

Haraway, Donna. 1988. 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective'. *Feminist Studies* 14 (3): 575-599.

Howard, Alexander Benjamin. 2014. *'The Art and Science of Data-Driven Journalism'*. New York, NY: Tow Centre for Digital Journalism, Columbia University. Available at https://academiccommons.columbia.edu/doi/10.7916/D8Q531V1.

Kitchin, Rob. 2021. *Data Lives: How Data Are Made and Shape Our World*. Bristol: Bristol University Press.

Larsen, Anna Grøndahl. 2017. 'Investigative Reporting in the Networked Media Environment: Journalists' Use of Social Media in Reporting Violent Extremism'. *Journalism Practice* 11 (10): 1231–45.

Leon, Sam. 2021. 'Accounting for Methods: Spreadsheets, Scripts and Programming Notebooks'. In *The Data Journalism Handbook*, edited by Liliana Bounegru and Jonathan Gray, 128–37. Amsterdam, The Netherlands: Amsterdam University Press.

Lowrey, Wilson, Ryan Broussard, and Lindsey A. Sherrill. 2019. 'Data Journalism and Black-Boxed Data Sets'. *Newspaper Research Journal* 40 (1): 69–82.

Melgar, Liliana, Marijn Koolen, Hugo Huurdeman, and Jaap Blom. 2017. 'A Process Model of Scholarly Media Annotation'. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval - CHIIR '17*, 305–8. Oslo, Norway: ACM Press.

Melgar-Estrada, Liliana, Marijn Koolen, Kaspar Beelen, Hugo Huurdeman, Mari Wigham, Carlos Martinez-Ortiz, Jaap Blom, and Roeland Ordelman. 2019. 'The CLARIAH Media Suite: A Hybrid Approach to System Design in the Humanities'. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 373–77. Glasgow Scotland UK: ACM.

Ordelman, Roeland, Liliana Melgar, Carlos Martinez-Ortiz, and Julia Noordegraaf. 2018. 'Media Suite: Unlocking Archives for Mixed Media Scholarly Research'. In *CLARIN Annual Conference 2018 Proceedings*, edited by Inguna Skadina and Maria Eskevich, 21–25. Pisa, Italy: CLARIN.

Parasie, Sylvain. 2015. 'Data-Driven Revelation?: Epistemological Tensions in Investigative Journalism in the Age of "Big Data"'. *Digital Journalism* 3 (3): 364–80.

Riche, Nathalie Henry, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. 2018. 'Introduction'. In *Data-Driven Storytelling*, edited by Nathalie

Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale, 1–15. Boca Raton, FL: CRC Press/Taylor & Francis Group.

Rinsdorf, Lars, and Raoul Boers. 2016. 'The Need to Reflect. Data Journalism as an Aspect of Disrupted Practice in Digital Journalism and in Journalism Education.' In *Proceedings of the International Association for Statistical Education.* Berlin: IASE. Avaialble at https://iase-web.org/documents/papers/rt2016/Rinsdorf.pdf.

Segel, E., and J. Heer. 2010. 'Narrative Visualization: Telling Stories with Data'. *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1139–48.

Stolper, Charles D., Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2016. 'Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories'. Available at https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/.

Van Gorp, Jasmijn, Liliana Melgar Estrada, and Julia Noordegraaf. 2021. 'Involving Users in Infrastructure Development: Methodological Reflections From the Research Pilot Projects Using the CLARIAH Media Suite'. *TMG Journal for Media History* 24 (1–2). Available at https://www.tmgonline.nl/article/10.18146/tmg.809.

Weber, Wibke. 2020. 'Exploring Narrativity in Data Visualization in Journalism'. In *Data Visualization in Society*, edited by Martin Engebretsen and Helen Kennedy, 295–311. Amsterdam: Amsterdam University Press.

Weber, Wibke, Martin Engebretsen, and Helen Kennedy. 2018. 'Data Stories. Rethinking Journalistic Storytelling in the Context of Data Journalism'. *Studies in Communication Sciences* 18 (1), 191–206.

Windhager, Florian, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. 2019. 'Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges'. *IEEE Transactions on Visualization and Computer Graphics* 25 (6): 2311–30.