Designing a Data Story: An Innovative Approach for the Selective Care of Qualitative and Ethnographic Data

Gaia Mosconi, Helena Karasti, Dave Randall and Volkmar Pipek

In this chapter, we present an explorative design concept for the sharing and reuse of qualitative-ethnographic data, that we call Data Story, which is inspired by data storytelling principles. Recent critics of data science have pointed to the need for a contextual approach to data, one which reflects the view that "data doesn't speak for itself, it needs a storyteller" (Duarte 2019, 5). However, approaches to data storytelling have hitherto mainly been contingent on the deployment and use of quantitative and statistical data. Our contribution suggests that considerable benefit might result from the use of new tools and methodologies inspired by data storytelling principles for qualitative data as well. We believe this approach has the potential to advance the Open Science agenda at large, which remains some way from realization, especially so for Humanities and Social Sciences (HSS) and for those researchers applying qualitative and ethnographic methods (Mosconi et al. 2019).

Policies that demand or encourage the release of data are predicated on the assumption that others will find the data useful and that data will thus be reused (Borgman 2012), but there is evidence indicating that secondary use of data is not yet an established practice (Borgman 2012; Bishop 2012, 2014; Mannheimer et al. 2019; Corti 2013). In our view, to make qualitative research data reusable means that, in addition to formats, (metadata) standards and licenses, we must pay attention to the practices of creating, structuring, analyzing, and interpreting data (Mosconi et al. 2019; Feger et al. 2020). In order to foreground this largely invisible work as a form of data care, we developed the concept of a Data Story and argue, along with Bellacasa (2011), that care is a useful conceptual anchor for this work specifically because it concerns itself with the "politics of knowledge." Caring is conceived of as entailing concern for the three dimensions of "labor/work, affect/affections, ethics/politics." Moreover, caring is interpreted as an act of doing and as a relational act of thinking-with data (Bellacasa, 2012). Our concept aligns with this insight, and in fact the Data Story supports collaborative mechanisms for

narration around data snippets that are situated at the center of its design. With it we propose the idea of data curation as an act of selective care that is foregrounded in the interface design.

The purpose of creating a Data Story is to provide a solution for the curation and sharing of data as it is expected by major funding agencies and institutions. In fact, this demand is seldom met in practice, and there aren't any tools available yet that clearly support this additional work of caring for the reusability of data (Mosconi et al. 2019). Therefore, with the Data Story concept, we wish to fill this gap. With our design, we aim to support researchers who do empirical work in organizing the data they care about and make explicit the context. In doing so, we hope to make easier the curation and sharing of qualitative and ethnographic data on the one hand, and the potential reuse by other researchers on the other hand. Software implementations of the Data Story concept will provide researchers with guides and templates supporting them to build stories around the most relevant data they have collected while at the same time envisioning a potential audience. We speculate on how this concept could potentially become a recognized publication format to be promoted in different collaborative data infrastructures or digital databases. In this way, researchers will have the opportunity to get recognition for this unrewarded and invisible work.

Our research concerns itself with the question: How can we best describe qualitative-ethnographic research data practices while respecting epistemological, methodological, and ethical challenges, in order to facilitate data sharing? Data Story, as an exploratory conceptual design solution, is an attempt to give an answer to this question. With it we wish to contribute to the international debate around Open Science, and encourage further engagement in such matters by scholars from various disciplines interested in the issues of openness and data care.

This chapter brings together various streams of literature on *Critical Data Studies* (Dalton and Thatcher 2014; Dalton, Taylor, and Thatcher 2016; Kitchin 2021), *data curation and sharing of qualitative-ethnographic work* (Bishop 2012, 2014; Corti 2013; Tsai et al. 2016; Treloar and Harboe-Ree 2008, Irwin 2013) and finally *data storytelling* (Duarte 2019, Knaflic 2015; Ojo, and Heravi 2018). Against the background of the outlined literature, we conducted empirical work and gained practical experiences within a research infrastructure project (INF) in which we engaged in formal and informal conversations with researchers working with qualitative-ethnographic data. Finally, we outline the exploratory design concept, Data Story, and discuss the act of selective care it affords.

Data as Matter of Care

As Dourish and Gómez have pointed out: "Data makes sense only to the extent that we have frames for making sense of it, and the difference between a productive data analysis and a random-number generator is a narrative account of the meaningfulness of their outputs" (2018, 8). The arrival of Big Data has been a motivating force for what is termed Critical Data Studies (Dalton and Thatcher 2014; Iliadis and Russo 2016). As Kitchin and Lauriault (2014) point out, critical data studies are largely concerned with questions about the nature of data: how they are being produced, organized, analyzed, and employed, and how best to make sense of them and the work they do, occasioned by a step change in the production and employment of data. The principal force of a critical approach, then, lies in the recognition that political, social, ethical, organizational, and economic elements shape data management as much as technical problems in much the way Bellacasa (2011) suggests in her critique of technoscience. As Bowker (2005) suggested:

> We need to open a discourse – where there is no effective discourse now – about the varying temporalities, spatialities and materialities that we might represent in our databases, with a view to designing for maximum flexibility and allowing as much as possible for an emergent polyphony and polychrony. Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care. (Bowker 2005, 184)

Thomer and Wickett (2020) further demonstrate the point through an analysis of the various material forms that the database can take, arguing that "'best practices' for data management are in tension with the realities and priorities of scientific data production," and "understanding pluralism in data practices is crucial to supporting the needs of those traditionally marginalized by information technologies—whether in their personal or disciplinary identity" (Thomer and Wickett 2020, 3). Curating for data work as a pluralistic and contextual endeavor has, as yet, not been fully realized.

Challenges for Qualitative Data Sharing

Data sharing and consequently data reuse have been extensively addressed (Heaton 2008; van den Berg 2008; Faniel and Jacobsen 2010). The vast part of the literature, however, deals with practices embedded in the natural and applied sciences. Our matter of care, however, is the additional complexity entailed in the management of qualitative data, where most of the challenges can be characterized as epistemological, methodological, and ethical in nature. For qualitative data, paying attention to the context of their collection and possible re-use becomes an overarching con-

cern. However, what context is, and how to describe it, is non-trivial (Moore 2006). Context determines whether something can be viewed as data or metadata and the "degree to which those contexts and meanings can be represented influences its transferability" (Borgman et al. 2018). Data loses meaning when removed from the original contexts, packaged in repositories, and disengaged from the knowledge and expertise of the researchers who performed the study (Walters 2009). When dealing with qualitative data we need to recognize the essentially reflexive character of data and that it is often rich with personal content (Tsai et al. 2016). Ethnographic approaches are generally based on a relationship of trust between researchers and participants, often in sensitive domains. This leads us to a consideration of the ethical challenges, where protecting the privacy of participants typically is one of the central aims (for more details see contribution by Kraus and Eberhard in this volume, and Eberhard and Kraus 2018).

Other challenges related to describing and preparing these types of data for sharing are: the lack of clear standards (Tsai et al. 2016; Antes et al. 2018) which are difficult to identify due to the heterogeneous nature and idiosyncrasy of researchers' data practices; not knowing how one might access and use the data in the future and for which purposes (Broom, Cheshire, and Emmison 2009); and finally time constraints where "the burden of organizing qualitative data for inspection or reuse could easily exceed the work of writing the manuscript itself" (Tsai et al. 2016, 5). As we shall see below, data storytelling provides us with inspiration as to how to best design for the curation and sharing of these types of data while addressing some of these complex issues.

Data Storytelling: Guiding Principles

The social sciences and humanities have long stressed the role that narrative plays in human life, in education and in research. As Game and Metcalfe argue:

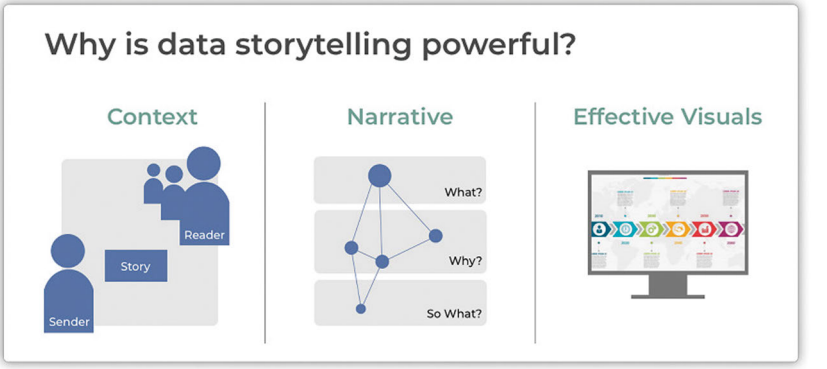
> Research is always an interpretative process that involves conversations and storytelling, though the research framework traditionally applies other names such as aims, methods and conclusions. Research conventions are a particular form of storytelling that allows sociologists and historians "to tell stories as if they weren't' storytellers." (1996, 65)

Social scientists tell stories for a range of different purposes. In doing so, they attempt to contextualize the "data" that they work with. They do so largely for analytic purposes. In relation to this, and to return to the question of what context is and how to describe it, there is a difference between context as an analytic construct – something that researchers, curators, etc. define – and something that emerges in and is enacted by the work of the participants. Put simply, context in this view has

no existence outside of the way in which it is ongoingly constructed by participants to an activity. Data, in other words, is a process of enactment. Digital storytelling, we want to argue, is a useful means to reconstruct what has previously been constructed or enacted.

Digital storytelling describes the practice of everyday professionals and organizations who make use of digital tools in order to tell a story. Digital stories can stimulate emotional responses in recipients and potentially offer interactive elements. Storytelling approaches have been applied to several fields: therapy, education, arts and culture, management and business, among others (Barrett 2006; De Vecchi et. al 2016; Yuksel 2011; Restrepo and Davis 2003; Denning 2006). In the last decade, however, due to the advent of Big Data and the "data revolution" (Kitchin 2014), western economies and governments are becoming progressively more data-driven, and therefore we have seen growing contributions and approaches focusing specifically on *Data Storytelling* (Duarte 2019; Knaflic 2015; Ojo and Heravi 2018). The main argument being made is that to understand and use data effectively, data needs to communicate a clear message (a narrative) and speak a human language to allow us to make sense of data (data sense making) and the reasons why it is presented (reconstructed) the way it is.

Figure 1: Main principles of Data Storytelling.



Source: https://www.nugit.co/what-is-data-storytelling/

As shown in the picture, three main principles summarize what data storytelling is about and how to achieve it: 1) explaining the context; 2) identifying a coherent narrative; 3) working on effective visualization. In data storytelling, the second principle, *narration*, is a crucial element. A narrative can, additionally, have emotional elements. A story has a beginning and an end, it has a goal, sometimes

a moral, and, obviously, a story has an audience. Narrative helps to "share norms and values, develop trust and commitment, share tacit knowledge, facilitate unlearning, and generate emotional connections" (Soule and Wilson 2002). The third principle is related to *effective visuals*. As Lee et al. (2015) suggest, relatively little attention has been paid in the visualization literature to the ways in which the stories in question are actually crafted.

To conclude, the concept of a Data Story for qualitative research data, as proposed here, combines all three affordances of data storytelling identified in the literature: a) it offers researchers an opportunity to provide contextual information to their collected data, b) it employs a narrative structure to demonstrate its analytical potential, c) and it allows for the integration of visual elements.

Background and Approach

Our research takes place in a research infrastructure project (INF), connected to the Collaborative Research Centre (CRC) "Media of Cooperation" funded by the DFG (in English: German Research Foundation) since January 2016 and is currently ongoing. Our CRC is characterized by interdisciplinary cooperation across disciplines and faculties, and most researchers apply qualitative and ethnographic methods. Being tasked with providing suitable solutions for both ongoing research and long-term preservation as well as the sharing of materials with a wider public, the focus of our project is on developing new RDM practices and infrastructures for qualitative-interpretative research contexts. Collaboration with the IT service provider of the University – a partner of the project – has been going on since the beginning of the funding period and this entailed interdisciplinary work with developers where we worked on metadata structures, restructured database hierarchies and classification schemes. Drawing on insights from CSCW and socio-informatics (Wulf et al. 2018), our project roots conceptual design and technology development itself in qualitative and long-term situated research. Therefore, we engaged in participatory observations, semi-structured interviews and informal conversations with CRC's projects, where we particularly investigated data practices, salient Research Data Management and data sharing issues that could inform our design.

The fieldwork we conducted as part of our infrastructural research was not straightforward and unproblematic. Some researchers felt annoyed and irritated by the work of our project. Its objectives were often met with indifference, questioned or overtly criticized on multiple occasions. In particular, *metadata critiques* emerged repeatedly during fieldwork. Researchers we talked to struggled to understand the meaning and the applicability of metadata standards such as the Dublin

Core[1] which was often mentioned by the IT service provider as the existing metadata standard that researchers should use to describe data for long-term preservation (and potentially for data reuse). However, in practice, qualitative researchers in particular lack familiarity with such standards and struggle to understand, or fail to see the point of, its technical language.

The agenda of the funding agency and the institutional top-down narratives around Research Data Management were not always matched by the immediate and practical objectives of research teams. Nonetheless, our approach was dialogic. Through interviews, observations and informal conversations we oriented reflexively to the often conflicting viewpoints expressed. We questioned design solutions, discussed current or new practices and the connection between the two in relation to design possibilities. As Schön (1983) pointed out, "design, in practice, is not a linear process." This pragmatic-reflexive approach led us to consider the need to embrace narrative as a focus for our deliberations in relation to data. The idea developed into what we call Data Story here which came about gradually after reflecting over a long period of time with local research groups. Their own narratives regarding data sharing and related challenges inspired the approach we describe. This led us to envision a system in which the showcasing of data snippets (or data nuggets) could potentially support the organization, curation and eventually sharing and reuse of research data, and therefore allow to meet the expectations of the funding body.

In the next section, we explain the major insights which led to the Data Story concept. We do so by grounding the concept in researchers' practices where storytelling emerges as an integral part of (collaborative) analytical work with qualitative data and therefore synergetic with these types of research approaches.

Grounding the Concept in Practices

The conceptualization of a Data Story gradually emerged during fieldwork, especially in our interaction via observations and interviews with researchers. Over three years, we paid particular attention to situations in which (informal) data sharing practices took place, and we observed how qualitative-ethnographic data was analyzed, collaboratively discussed, and represented with the support of (digital) media.

We began to notice, for instance, the common practice in qualitative research of sharing *data snippets* in collaborative analysis sessions with members of the same project (but with different disciplinary backgrounds) and/or with researchers from other projects. In these situations, snippets of anonymized data are often selected,

1 https://www.dublincore.org/specifications/dublin-core/dces/.

enriched with context and sent to participants via email a few days before the analysis session. A *narration* or, if you will, a *story* which contextualizes the data is often provided by the data collector in written form (i.e. as text), and/or in oral form at the very beginning of the session. The piece of data in question then is often projected in the room in order to guide the conversation and to promote interpretative work. Through this collaborative practice, as Dourish and Cruz (2018) expressed it, data is "put to work in particular contexts, sunk into narratives that give them shape and meaning, and mobilized as part of broader processes of interpretation and meaning-making" (Dourish and Cruz 2018, 1). Data are not collected and analyzed in a vacuum, but are always shaped, co-created, (partially) shared and narrated based on the specific circumstances in which data are needed and "put to work." Another example is Rose, who said: "in our team we couldn't really do very close readings of the data together, due to lack of time and the overload of data we collected, so we just selected a few data and sketches that we could talk about in order to collaboratively develop our thinking." Her team developed "ad hoc" visualization techniques around data snippets, as we might call it, in order to elicit a collaborative narrative and which partly inspired our conceptual design.

Another researcher, Sophie, told us that direct access to data (even if partial) could foster interdisciplinary collaboration and new research approaches: "sometimes you see a paper, but you do not realize all the kinds of data and fieldwork that has been done, and if you look at the data then it makes you think of other collaboration that you could have with this person." In fact, Sophie had collaborated with a social scientist in the past, but only after looking at some examples of ethnographic data was she capable of understanding what kinds of collaboration might be possible and what research questions could be answered. But she also added that "there aren't really good solutions to represent and share ethnographic data just yet" and "we had to share the data via email which obviously wasn't ideal!" Another important element connected to data sharing and reuse is the messiness of ethnographic work. The majority of researchers we talked to expressed discomfort in sharing their qualitative data due to the "messiness" which often comes with it. We noticed their need to have better tools and techniques that could support the organization of the heterogenous data and the non-linear way of conducting research typical of ethnographic work. The Data Story started to emerge then as a form of digital data storyboarding to support collection, organization, collaboration, and data sense-making.

The above vignettes point to the way in which a *storytelling approach* to data curation can be called into action, one which is more aligned with researchers' practices, and as possible inspiration to organize the heterogeneous data and to support collaborative data sense-making. In the following, we demonstrate how the Data Story is envisaged to work by showing the design sketches of the low-

fidelity prototype we have developed so far. We will then discuss more extensively the idea of selective care that it affords.

The Data Story Process and its Components

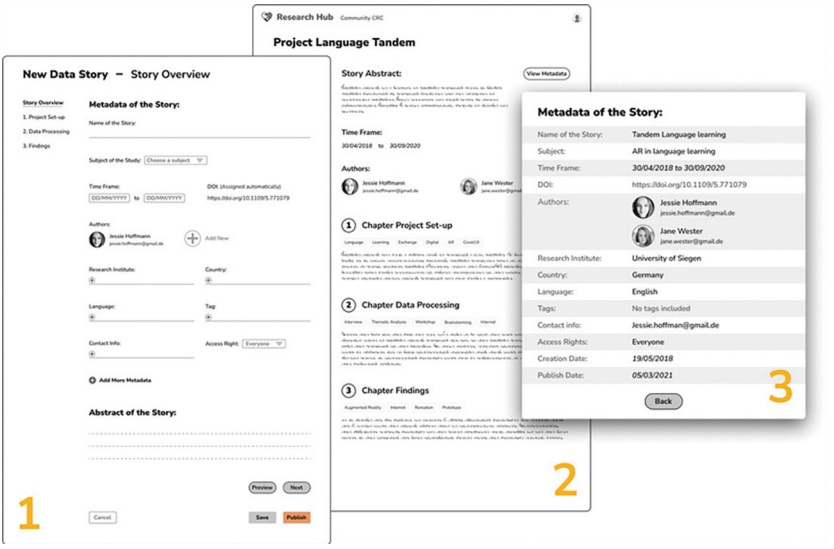
This design concept is meant to be an organizing device in support of (collaborative) storytelling practices as a major component of data analysis and sense making. By engaging with its process and its interactive interface researchers will have the opportunity to perform data curation practices resulting in selected data snippets. In this way, we wish to make easier the sharing of these types of data on the one hand, and the potential reuse by external researchers on the other hand.

The interface is organized into chapters to sort the shared data into sections and better help in navigating through the story. The chapters sequence creates a timeline of the actions, events, and decisions regarding the study being shared. Each chapter might have multiple data snippets that help clarify the overall story. Questions and tips are highlighted in the interface of each chapter to support reflexivity, elicit discussions and help researchers to construct their narrative. To exemplify the possibilities, we provide a possible structure with an initial overview screen (0) followed by three main chapters for the story: (1) project set-up; (2) data processing (with snippets of anonymized data), and (3) main findings. As mentioned before, each chapter provides a focused insight into the study conducted but also it invites to make explicit the context and to define a coherent narrative.

(0) Overview Screen

In the overview screen, general information regarding the study will be given, like the time frame and to which project it belongs (a single publication, a complete research project, a PhD dissertation, and so on). Moreover, the authors can introduce themselves, their research institution, their contact information, etc. This is needed to connect a Data Story with a specific researcher or research team (in order to be publicly acknowledged, and possibly contacted).

Figure 2: Data Story module overview: Figure 2.1 is the view of the author, 2.2 is the view of the reader, and 2.3 is an overview of some of the included metadata



(1) The Project Set-Up Chapter

The project set-up chapter introduces the overall story outline, in order to provide an understandable context for the study. Information related to the research field, topic, and research questions of the study, as well as methods used, a short summary about the motivation and aim of the study can be included. Tips and questions are highlighted in the interface in order to elicit reflexive thinking while supporting data sense-making.

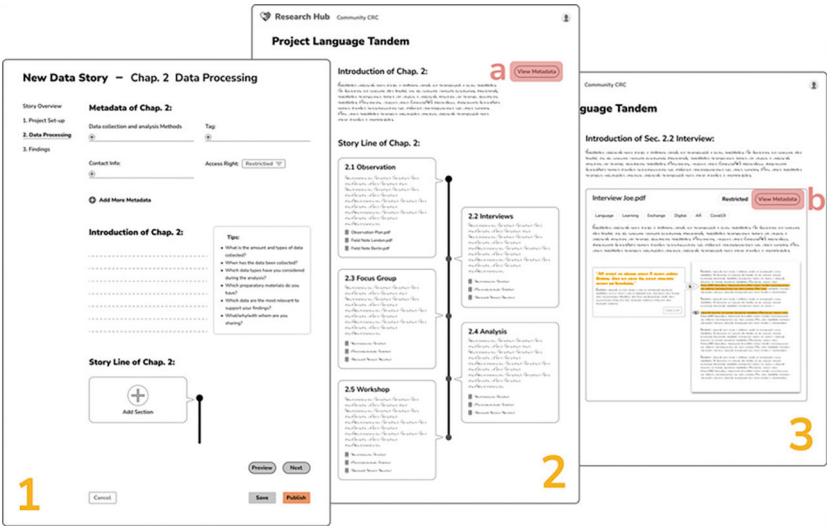
(2) The Data Processing Chapter

The data processing chapter encapsulates the actual *data snippets*. It also provides a more detailed contextual narrative that explains important milestones in the data collection and the analysis process. As with the project set-up chapter, the process narrative is aimed at resolving common queries to support the sense making of the shared data nuggets.

The chapter provides the possibility to create sub-sections which categorize and group data, based on the data type, to ease navigation through it. It is advised to create and fill the sub-sections with relevant data in a way that supports the

storyline and sequence of the story. Moreover, this chapter creates a storyline by ordering the created sub-sections sequentially. Authors of the stories will have the ability to relocate the created sub-section if necessary by dragging it to the desired location on the storyline.

Figure 3: Data processing chapter: Figure 3.1 shows the view of the story writer, 3.2 shows the story from the reader's view after publishing, 3.3 shows the interview sub-section.



The Data Story supports the sharing of different data formats. Some snippets might be extracted from a text file and thus have a text format, e.g., interview questions, transcripts, notes, etc. Other data snippets might take the shape of audio or video files, presentations, posters, pictures, sketches, and design material, etc. As in the chapter before, the author will be provided with a list of questions that might add a better structure to the story and support the sense making of the shared data as well as enrich the contextual layer.

As already mentioned, only selected and anonymized data will be displayed. This is for three reasons: (1) facilitate the protection of the study participants and avoid the disclosure of any private and sensitive information; (2) decrease data overload by encouraging researchers to display only the most relevant pieces of data; (3) time constraints: as it is not possible to provide a deliberate narrative, in a relatively short time, that is rich of context to all the collected data of the study.

(3) The Findings Chapter

Last but not least is the Findings chapter, where the narrative is brought to an end and future visions can be explained. Any publications or material, citation and review data can be included in this chapter. Again, guiding questions and tips for contextualizing the chapter will be visible upfront and will help researchers in structuring the information and narrative.

Supporting Processual Workflows: Plugin Solution

The Data Story aims to promote curation activities to be carried out as soon as possible, as close as possible to the data source, and in support of workflow. It is a proposal for embeddedness. In order to achieve this, the Data Story will be connected to tools used routinely while collecting, analyzing and processing data. Therefore, a plugin solution is envisioned. The plugin is to be connected to text editing software like Microsoft Word, data analysis tools like MaxQDA, literature management tools like Zotero, cloud storage tools like Sciebo[2] or other tools that researchers routinely use. As mentioned earlier, the idea is to provide the researchers the opportunity to feed their Data Story with data at all times by creating such direct connections between a collaborative research infrastructure already in use and the researcher's data storage. In other words, researchers can select key data pieces (text, file, etc.) while organizing and analyzing their data, and send them to the Data story as *data snippets*. Moreover, researchers will be given the chance to add annotations, descriptions, comments, and metadata that clarify the context of the chosen data. The transferred data snippets can be previewed and further annotated via the interface.

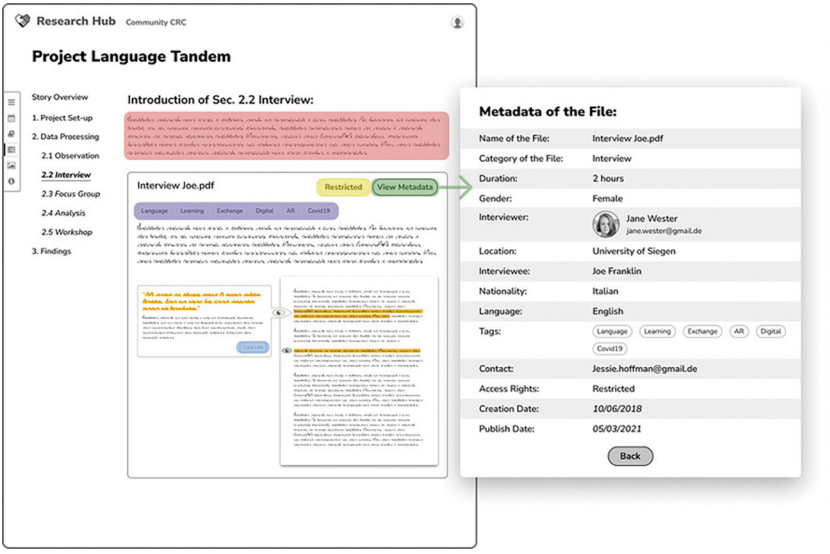
Publishing: DOI and Accessibility Rights

Once researchers have completed their Data Story, and feel secure with the provided data and narrative, they will be able to publish it. A DOI (Digital Object Identifier) can also be (automatically) assigned to the Data Story (see figure 4, blue highlight). We envision a new practice that could emerge from this: the DOI link of the Data Story web-interface might be promoted in papers where potential collaborators or interested parties could see additional data. Moreover, share links will be (automatically) generated for single data entries to indicate a clear reference to a specific data snippet.

2 Info on Sciebo: https://hochschulcloud.nrw/en/index.html.

Researchers can share parts of the data with some recipients and some other parts with some other audience using the same Data Story. This is facilitated by different accessibility rights provided in the Data Story for each data snippet added in the storyline. Taking inspiration from Jones et al. (2018) we considered the following accessibility rights: open, restricted, controlled, and closed (these categories can be assigned to the whole Data Story, or to specific data snippets). The accessibility right *Open* means that data is available to be accessed by anyone; *Restricted* means to be accessed by some specific audience; *Controlled*, means that the author has to grant permission to access it after assessing the request. Lastly, *Closed* means "data deposit and citation exist for archival purposes but no data are currently available (could be embargoed until publication of results, change in sensitive situation, death of a participant, or certain duration of time from collection)" (Jones et al. 2018, 21). Figure 4 highlights how accessibility rights will be shown in the design (highlighted in yellow).

Figure 4: Visual of metadata, tags, DOI, data snippet and the story (Purple: tags, Red: Story. Orange: Data Snippets, Blue: DOI, Green: Metadata, Yellow: access rights)



In our view, the Data Story should be promoted as a new publication format that is centered around relevant data points. Data Stories can act as an intermediate format between a larger dataset to be stored and secured in long-term archives and the official publications (paper, books etc.). Data Story could offer insights into the content of a dataset but also offering some reflections on the data that might

not be part of the final publications. By promoting a Data Story as a new publication format that can be cited, researchers will have the incentives to actually engage with this type of work and get rewarded for this additional effort. For being a Data Story an additional step is important so that researchers will get compensated for this work. By envisioning an accessible open link, Data Stories can circulate freely through the web, and can act as entry points for engagement with the data that have been collected.

We are planning to implement this design in a collaborative research infrastructure, called Research-hub, that is already in use in our research center. However, we believe this design with its modular and customizable characteristics has the potential to be integrated as interface layer of any other (collaborative) data infrastructure or digital database.

Data Story as an Act of Selective Care

Above, we have described an approach, inspired by storytelling insights and designed to support a workflow for the organization, curation and sharing of data which can be used in conjunction with more standard approaches and data descriptions (i.e. metadata). The purpose of creating the Data Story is to provide all those with an interest in the possible uses of data with an easy way to access and understand how a data collection was assembled and the reasons for it. This, we do by supporting researchers who collected the data in the first place to envision a possible audience and to make the context of their work explicit, using both metadata and a narrative. So, this design concept is meant to be an organizing device in support of (collaborative) storytelling practices as a major component of data analysis and sense making. As we have seen, however, complex issues intervene. They include the nature of the work, ethical concerns, and the reflexive nature of the engagement with data, all of which have methodological and epistemological consequences.

We take on board the injunction of van Es and Schäfer that, "[r]ather than import questions and methods from the hard sciences, we must develop our own approaches and sensitivities in working with data that will reflect the humanities' traditions" (2017, 16). The authors here include a call for action, inviting humanities scholars to develop their own research questions and methods to stay consistent with their epistemological positions. We have shown how we might translate these ideas to the field of Research Data Management and curation. If solutions to data sharing and curation need to be found, as expected and demanded by funding agencies, then we argue, those technical solutions, tools or infrastructures will need to embrace and embed in the design cultural values, methodological practices and epistemological understandings of the communities they are designed for. In

doing so, we again connect to the concept of care as pushed forward by Bellacasa (2011): "… representing matters of fact and sociotechnical assemblages as matters of care is to intervene in the articulation of ethically and politically demanding issues. The point is not only to expose or reveal invisible labors of care, but also to generate care" (Bellacasa 2011, 94). We discuss below two lines of argument in which we explicit how the Data Story reveals the invisible labor of data care while at the same time generating care for both the data producer and the data re-user.

Complementing Metadata Standards with a Story

As we have seen, it is now accepted that context is critical to our understanding of data (Borgman 2015; Carlson and Anderson, 2007) as a representational mechanism bridging data producers and data re-users. Within the Research Data Management domain this contextual role is typically assigned to metadata standards and data descriptions. Formal and standardized metadata such as the Dublin Core or the Data Documentation Initiative (DDI) assume not only a contextual role but also, it is claimed, are essential for the discovery, comprehension, and reuse of data. Metadata are often interpreted as the "bridges" because they can, in principle, convey the information essential for discovery and secondary analysis: "secondary users must rely on the amount of formal metadata that travels along with the data in order to exploit their full potential." (Ryssevik, 2021). However, and as is evidenced both in our own practical experiences with researchers and in Feger et al. (2020), cleaning the data, and filling metadata requirements is a quite tedious and rather technical practice. The inherent difficulties, along with the fact that researchers do not see this as their primary purpose, means it is frequently poorly done or not done at all. Moreover, analysts of qualitative data often do not have enough time to fully explore their data given the richness and the amount of the data in question (Fielding and Fielding, 2000; Yoon 2014). Therefore, the Data Story provides the opportunity to display only selected data snippets and narrate them coherently. This we argue could potentially make it easier for a researcher interested in certain data sets to understand how the data collection and analysis came about. At the same time, the researcher(s) who collected the data is supported in explaining the whole data process, displaying what, for them, is the most important aspect in the data and envisioning a potential re-user.

The Data Story interface makes visible the act of care by articulating the tasks of data care needed in order to organize the data, retrieve them, present them, share them, and possibly reuse them. In fact, it provides every chapter with the option to annotate, tag and add metadata. The Data Story suggests metadata (i.e.: the Dublin Core or DDI) as the standards source for elements set. They can, however, be adapted quickly and added as new folksonomy. In this way, metadata are

treated as "living things" that can grow and develop based on a bottom-up understanding. As mentioned earlier, the Data Story invests noticeable effort in bringing the data and its metadata together by integrating many of the important metadata fields in its interface in a way that makes metadata an important pillar of the story narrative and driver of discussions. It promotes data literacy and awareness, as it is an opportunity for researchers to learn about the role of metadata but also put it into question and adapt it to their needs.

With our contribution, we complement the role that formal data descriptions (metadata) bring to the table when they are provided, and suggest an alternative when they are not, depending on the institutional investment in data curation. By focusing on narrative as an organizational layer and as a useful method to make explicit the context, we aim to make the interpretative work – essential to make use of data – less onerous for both parties: data producer(s) and data re-user(s). Stories, then, can serve a further purpose, that of inviting re-users to reflect on what messages can be found in the data, what questions can be evoked and answered, and what uses the data can be put to. The Data Story is then a complementary organizing layer – flexible, culturally, collaborative and context sensitive – that can be added to the formal and structured way of organizing and preserving data. Finally, by promoting the Data Story as a possible intermediate publication format, we allow researchers to get rewarded for this additional step and we show care for their additional curation work.

Designing for Situated Data

That knowledge is situated is hardly a discovery by now and, indeed, has been a central tenet of the sociology of knowledge at least since Mannheim (1936). It can be traced through the work of, for instance, Vygotsky (1980), Garfinkel (1967) and many others, but has been reinvigorated in practice-oriented thinking (see e.g., Randall et al. 2018) and in feminist standpoint theory (Haraway 1991; D'Ignazio and Klein 2020). Critical Data Studies (Dalton and Thatcher 2014; Dalton, Taylor, and Thatcher 2016; Kitchin 2021) draws on these insights to address "the situated, partial, and constitutive character of knowledge production" (Drucker 2011, 2), in order to show how the meaning of data is derived from its context of production and use. This is particularly true for qualitative data because qualitative research is characterized as an "insider activity" (Mauthner et al. 1998), its knowledge "is highly contextual and experience dependent" (Niu and Hedstrom 2008), its practice uses "the self … as the primary instrument of knowing" (Ortner 2006), and it involves interpretation and subjectivities not concrete (or transportable) enough for information to be documented and reused in its entirety (Broom, Cheshire, and Emmison 2009).

Kitchin (2021) suggests that, for *all* datasets, "we tell stories *about* data, and stories *with* data, in which there are inherent politics at play in how they are discursively figured" (Kitchin 2021, 5). D'Ignazio and Klein in their book *Data Feminism* (2020) also pose interesting questions such as, "How can we use data to remake the world? […] or, more precisely, whose information needs to become data before it can be considered as fact and acted upon?" (D'Ignazio and Klein 2020, 36). Embracing the partiality and situatedness of data means designing with these questions in mind, to question what is data, what is metadata, how do we construct facts and information, how are they disseminated, how they get curated and shared. In this way, the Data Story concept engages in "politics of knowledge" (Bellacasa 2011). Our design helps to address the questions raised above and tries to give some answers applied to the context of curation and data sharing. With our design, we wish to support pluralism in research data (management) practices, embrace situated knowledge, without excluding data collection efforts which might not fit neatly into current standards and categories.

Concerning the issue of reuse, the question is how does the Data Story provide a narrative which can not only contextualize the production of the data but also render it relevant for the re-user. Of course, there is not, and cannot be, any simple answer to such a question, for the value of data in reuse will depend as much on the reasons for reuse as it does on the reasons for its production. Nevertheless, the Data Story can do a number of useful things (bearing in mind that it is a complement to, and not a replacement for, established metadata schemes). Firstly, and most obviously, it renders certain features of the data more visible which otherwise would not be (at least immediately) the case. The proposed three-chapters structure affords a number of data relevancies and highlight specific data points. Thus, the project set-up might tell the re-user why the data exists in the first place, what value it is believed to add to existing knowledge, information about the disciplinary origins of researchers (and possibly the backgrounds of participants). The data processing section affords snippets which go some way to answering the queries that re-users may have about methods adopted, the amount of data and its formats, examples of the data in question, and so on. The findings section provides a link from the snippets to results, enabling judgements about accuracy, reliability and validity to be made, literature deemed to be relevant to the researchers, reviews of the work, and so on. Overall, it offers the possibility of comparison with the aims that re-users might have, the options they may have with regard to methods and forms of analysis, insights into the kinds of questions and answers embedded in the data, insights into the number and type of people they may wish to engage with, and even suggest options for future progress.

Conclusion

Organizing, communicating, and understanding data are crucial issues of our modern "datafied society" (van Es and Schäfer 2017). Yet, in our digital world it is not always clear what data are, how best to make sense of them, and what is at stake (Kitchin 2021, 1). With our design concept of the Data Story, we aim at fostering exchange around data storytelling which should not be limited to quantitative data, data visualizations, infographics, statistics, and standard approaches, but should embrace a plurality of data practices and approaches.

Bellacasa (2012) argued: "We cannot possibly care for everything, not everything can count in a world, not everything is relevant in a world…" (Bellacasa 2012, 204). For this reason, the Data Story aims at showcasing only anonymized data snippets (such as interview excerpts, pictures, videos, sketches or any other relevant material) that researchers are encouraged to select based on the relevance for their own research findings and for an envisioned audience i.e. what they care about. This act of selective care is organized along a timeline and enhanced with storytelling practices (in oral and written form). STS scholars have already demonstrated how formal data descriptions wrapped in informal descriptions might increase the usefulness of the data (Bowker and Star 1999). The Data Story concept embraces this insight. In fact, it integrates traditional metadata standards but also allows the creation of bottom-up folksonomies. Metadata elements, folksonomy and data snippets are then visualized and glued together, enriched and situated with the addition of a storyline. In this way, Data Story brings the invisible work of data care to the forefront, it promotes data awareness and reflexivity, and calls for making visible (and supporting) curation activities, its concerns, technicalities, and specificities while articulating workflows and processes for collaborative activities. In all, the notion of care and more specifically how selective caring (or caring about caring) provide a conceptual anchor for a range of issues that have hitherto been only addressed in very limited ways. The Data Story, we suggest, is an explorable avenue for more sophisticated approaches to data management and reuse.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 262513311 – SFB 1187 "Media of Cooperation".

References

Antes, Alison L., Heidi A. Walsh, Michelle Strait, Cynthia R. Hudson-Vitale, and James M. DuBois. 2018. "Examining Data Repository Guidelines for Qualitative Data Sharing." *Journal of Empirical Research on Human Research Ethics* 13 (1): 61–73.

Barrett, Helen. 2006. "Researching and Evaluating Digital Storytelling as a Deep Learning Tool." In Proceedings of SITE 2006, edited by C. Crawford, R. Carlsen, K. McFerrin, J.Price, R. Weber and D.Willis, 647–54. Orlando: Association for the Advancement of Computing in Education (AACE). Accessed September 8, 2021. https://www.learntechlib.org/primary/p/22117/.

Bellacasa, Maria Puig de la. 2011. "Matters of Care in Technoscience: Assembling Neglected Things." *Social Studies of Science* 41 (1): 85–106. Accessed September 8, 2021. https://doi.org/10.1177/0306312710380301.

Bellacasa, María Puig de la. 2012. "'Nothing Comes Without Its World': Thinking with Care." *The Sociological Review* 60 (2): 197–216. Accessed September 8, 2021. https://doi.org/10.1111%2Fj.1467-954X.2012.02070.x

Berg, Harry van den. 2008. "Reanalyzing Qualitative Interviews from Different Angles: The Risk of Decontextualization and Other Problems of Sharing Qualitative Data." *Historical Social Research* 33 (3): 179–92. Accessed September 8, 2021. https://doi.org/10.12759/hsr.33.2008.3.179-192.

Berry, David M. 2012. "Introduction: Understanding the Digital Humanities." In *Understanding Digital Humanities*, edited by David M. Berry, 1–20. New York: Palgrave Macmillan US. Accessed September 8, 2021. https://doi.org/10.1057/9780230371934_1.

Bishop, Libby. 2012. "Using Archived Qualitative Data for Teaching: Practical and Ethical Considerations." *International Journal of Social Research Methodology* 15 (4): 341–50. Accessed September 8, 2021. https://doi.org/10.1080/13645579.2012.688335.

Bishop, Libby. 2014. "Re-Using Qualitative Data: A Little Evidence, on-Going Issues and Modest Reflections." *Studia Socjologiczne* 3 (214): 167–76.

Blackwell, Lindsay, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. "Classification and Its Consequences for Online Harassment: Design Insights from HeartMob." *Proceedings of the ACM on Human-Computer Interaction* 1 (CSCW): 1–19. Accessed September 8, 2021. https://doi.org/10.1145/3134659.

Borgman, Christine L. 2012. "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology* 63 (6): 1059–78. Accessed September 8, 2021. https://doi.org/10.1002/asi.22634.

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Borgman, Christine L., Andrea Scharnhorst, and Milena S. Golshan. 2018. "Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse." *ArXiv:1802.02689 [Cs]* 70 (November). Accessed September 8, 2021. https://arxiv.org/abs/1802.02689.

Bowker, Geoffrey C. 2005. *Memory Practices in the Sciences*. Cambridge: MIT Press.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Braidotti, Rosi. 2013. *The Posthuman*. Cambridge, UK/Malden, MA: Polity Press.

Broom, Alex, Lynda Cheshire, and Michael Emmison. 2009. "Qualitative Researchers' Understandings of Their Practice and the Implications for Data Archiving and Sharing." *Sociology* 43 (6): 1163–80. Accessed September 8, 2021. https://doi.org/10.1177%2F0038038509345704.

Bullard, Julia. 2014. "Values and Negotiation in Classification Work." In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing – CSCW Companion '14*, February 2014: 45–48. Accessed September 08, 2021. https://dl.acm.org/doi/10.1145/2556420.2556820.

Carlson, Samuelle, and Ben Anderson. 2007. "What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use." *Journal of Computer-Mediated Communication* 12 (2): 635–51. Accessed September 8, 2021. https://doi.org/10.1111/j.1083-6101.2007.00342.x.

Corti, Louise. 2013. "Infrastructures for Qualitative Data Archiving." In *Forschungsinfrastrukturen für die Qualitative Sozialforschung*, edited by Denis Huschka, Hubert Knoblauch, Claudia Oellers, and Heike Solga, 35–62. Berlin: Scivero. Accessed September 8, 2021. https://www.konsortswd.de/wp-content/uploads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.

Craig M. Dalton and Jim Thatcher. 2014. "What Does A Critical Data Studies Look Like, And Why Do We Care?" *Society and Space*. Accessed September 8, 2021. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care.

Dalton, Craig M, Linnet Taylor, and Jim Thatcher. 2016. "Critical Data Studies: A Dialog on Data and Space." *Big Data & Society* 3 (1): 1–9. Accessed September 8, 2021. https://doi.org/10.1177%2F2053951716648346.

De Vecchi, Nadia, Amanda Kenny, Virginia Dickson-Swift, and Susan Kidd. 2016. "How Digital Storytelling Is Used in Mental Health: A Scoping Review." *International Journal of Mental Health Nursing* 25 (3): 183–93. Accessed September 8, 2021. https://doi.org/10.1111/inm.12206

Denning, Stephen. 2006. "Effective Storytelling: Strategic Business Narrative Techniques." *Strategy & Leadership* 34 (1): 42–48. Accessed September 8, 2021. https://doi.org/10.1108/10878570610637885.

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

Dourish, Paul, and Edgar Gómez Cruz. 2018. "Datafication and Data Fiction: Narrating Data and Narrating with Data." *Big Data & Society* 5 (2): 1–10. Accessed September 8, 2021. https https://doi.org/10.1177%2F2053951718784083.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5 (1): 1–21.

Duarte, Nancy. 2019. *Data Story: Explain Data and Inspire Action through Story*. Oakton, Virginia: Ideapress Publishing.

Eberhard, Igor, and Wolfgang Kraus. 2018. "Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien." *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (1): 41–52.

Es, Karin van, and Mirko Tobias Schäfer, eds. 2017. *The Datafied Society: Studying Culture through Data*. Amsterdam: Amsterdam University Press.

Faniel, Ixchel M., and Trond E. Jacobsen. 2010. "Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data." *Computer Supported Cooperative Work* 19 (3–4): 355–75. Accessed September 8, 2021. https://doi.org/10.1007/s10606-010-9117-8

Feger, Sebastian S., Paweł W. Wozniak, Lars Lischke, and Albrecht Schmidt. 2020. "'Yes, I Comply!': Motivations and Practices around Research Data Management and Reuse across Scientific Fields." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2): 1–26. Accessed September 8, 2021. https://doi.org/10.1145/3415212.

Fielding, Nigel G., and Jane L. Fielding. 2000. "Resistance and Adaptation to Criminal Identity: Using Secondary Analysis to Evaluate Classic Studies of Crime and Deviance." *Sociology* 34 (4): 671–89. Accessed September 8, 2021. https://doi.org/10.1177%2FS0038038500000419

Game, Ann, and Andrew W. Metcalfe. 1996. *Passionate Sociology*. London/Thousand Oaks/ New Delhi: SAGE Publications.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.

Haraway, Donna Jeanne. 1991. *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Heaton, Janet. 2008. "Secondary Analysis of Qualitative Data: An Overview." *Historical Social Research / Historische Sozialforschung* 33 (3): 33–45.

Iliadis, Andrew, and Federica Russo. 2016. "Critical Data Studies: An Introduction." *Big Data & Society* 3 (2): 1–7.

Irwin, Sarah. 2013. "Qualitative Secondary Data Analysis: Ethics, Epistemology and Context." *Progress in Development Studies* 13 (4): 295–306. Accessed September 8, 2021. https://doi.org/10.1177%2F1464993413490479.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London/Thousand Oaks/ New Delhi: SAGE Publications. Accessed September 8, 2021. https://dx.doi.org/10.4135/9781473909472.

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

Kitchin, Rob, and Tracey Lauriault. 2014. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." *SSRN Scholarly Paper*. Accessed September 8, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112.

Knaflic, Cole Nussbaumer. 2015. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: John Wiley & Sons.

Lee, Bongshin, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. "More Than Telling a Story: Transforming Data into Visually Shared Stories." *IEEE Computer Graphics and Applications* 35 (5): 84–90. Accessed September 8, 2021. https://doi.org/10.1109/MCG.2015.99.

Mannheim, Karl. 1936. *Ideology and Utopia*. London: Routledge.

Mannheimer, Sara, Amy Pienta, Dessislava Kirilova, Colin Elman, and Amber Wutich. 2019. "Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges." *American Behavioral Scientist* 63 (5): 643–64. Accessed September 8, 2021. https://doi.org/10.1177%2F0002764218784991.

Martin, David, Jacki O'neill, Dave Randall, and Mark Rouncefield. 2007. "How Can I Help You? Call Centres, Classification Work and Coordination." *Computer Supported Cooperative Work* 16 (3): 231–64. Accessed September 8, 2021. https://doi.org/10.1007/s10606-007-9045-4.

Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. "The Data Are Out There, or Are They? Implications for Archiving and Revisiting Qualitative Data." *Sociology* 32 (4): 733–45. Accessed September 8, 2021. https://doi.org/10.1177%2F0038038598032004006.

Moore, Niamh. 2006. "The Contexts of Context: Broadening Perspectives in the (Re)Use of Qualitative Data." *Methodological Innovations Online* 1 (2): 21–32. Accessed September 8, 2021. https://doi.org/10.4256%2Fmio.2006.0009.

Mosconi, Gaia, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. "Three Gaps in Opening Science." *Computer Supported Cooperative Work* 28 (3): 749–89. Accessed September 8, 2021. https://doi.org/10.1007/s10606-019-09354-z.

Niu, Jinfang, and Margaret Hedstrom. 2008. "Documentation Evaluation Model for Social Science Data." *Proceedings of the American Society for Information Science and Technology* 45 (1): 11–11. Accessed September 8, 2021. https://doi.org/10.1002/meet.2008.1450450223.

Ojo, Adegboyega, and Bahareh Heravi. 2018. "Patterns in Award Winning Data Storytelling." *Digital Journalism* 6 (6): 693–718. Accessed September 8, 2021. https://doi.org/10.1080/21670811.2017.1403291.

Ortner, Sherry B. 2006. *Anthropology and Social Theory: Culture, Power, and the Acting Subject*. Durham, NC: Duke University Press. Accessed September 8, 2021. https://doi.org/10.1215/9780822388456.

Randall, Dave, Markus Rohde, Kjeld Schmidt, and Volker Wulf, eds. 2018. "Introduction: Socio-Informatics— Practice Makes Perfect?" In *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. Oxford: Oxford Uni-

versity Press. Accessed September 8, 2021. https://doi.org/10.1093/oso/9780198733249.003.0001.

Restrepo, Elizabeth and Lisa Davis. 2003. "Storytelling: Both Art and Therapeutic Practice." *International Journal of Human Caring* 7 (1): 43–48. Accessed September 8, 2021. https://doi.org/10.20467/1091-5710.7.1.43.

Ryssevik, Jostein. 2021. "The Data Documentation Initiative (DDI) metadata specification." *DDI*. Accessed September 8, 2021. https://ddialliance.org/sites/default/files/ryssevik_0.pdf.

Schön, Donald. A. 1983. *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Soule, Deborah, and Daniel Gray Wilson. 2002. "Storytelling in Organizations: The Power and Traps of Using Stories to Share Knowledge in Organizations." *LILA* 53 (January): 1–12.

Thomer, Andrea K., and Karen M. Wickett. 2020. "Relational Data Paradigms: What Do We Learn by Taking the Materiality of Databases Seriously?" *Big Data & Society* 7 (1): 1–16. Accessed September 8, 2021. https://doi.org/10.1177%2F2053951720934838.

Treloar, Andrew, and Cathrine Harboe-Ree. 2008. "Data management and the curation continuum: How the Monash experience is informing repository relationships." Lecture VALA 2008: The 14th Biennial Conference and Exhibition, 5 – 7 February 5–7, 2008, Melbourne.

Tsai, Alexander C., Brandon A. Kohrt, Lynn T. Matthews, Theresa S. Betancourt, Jooyoung K. Lee, Andrew V. Papachristos, Sheri D. Weiser, and Shari L. Dworkin. 2016. "Promises and Pitfalls of Data Sharing in Qualitative Research." *Social Science & Medicine* 169 (November): 191–98. Accessed September 8, 2021. https://doi.org/10.1016/j.socscimed.2016.08.004.

Vygotsky, Lev. S. 1978. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. Accessed September 8, 2021. https://doi.org/10.2307/j.ctvjf9vz4.

Walters, Peter. 2009. "Qualitative Archiving: Engaging with Epistemological Misgivings." *Australian Journal of Social Issues* 44 (3): 309–20. Accessed September 8, 2021. https://doi.org/10.1002/j.1839-4655.2009.tb00148.x

Wulf, Volker, Volkmar Pipek, David Randall, Markus Rohde, Kjeld Schmidt, and Gunnar Stevens, eds. 2018. *Socio-informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. Oxford, UK: Oxford University Press.

Yoon, Ayoung. 2014. "'Making a Square Fit into a Circle': Researchers' Experiences Reusing Qualitative Data." *Proceedings of the American Society for Information Science and Technology* 51 (1): 1–4. Accessed September 8, 2021. https://doi.org/10.1002/meet.2014.14505101140.

Yuksel, Pelin, Bernard Robin, and Sara McNeil. 2011. "Educational Uses of Digital Storytelling All around the World." International Conference Society for Infor-

mation Technology & Teacher, March 07, 2011, Nashville, Tennessee,1264–71. Waynesville, NC: Association for the Advancement of Computing in Education. Accessed September 8, 2021. https://www.learntechlib.org/primary/p/36461/.