Designing a Data Story: A Storytelling Approach to Curation, Sharing and Data Reuse in Support of Ethnographically-driven Research

GAIA MOSCONI, University of Siegen, Germany

DAVE RANDALL, University of Siegen, Germany

HELENA KARASTI, IT University of Copenhagen, Copenhagen

SAJA ALJUNEIDI, OFFIS - Institute for Information Technology, Germany

TONG YU, University of Siegen, Germany

PETER TOLMIE, University of Siegen, Germany

VOLKMAR PIPEK, University of Siegen, Germany

In this paper, we introduce an innovative design concept for the curation of data, which we call 'Data Story'. We view this as an additional resource for data curation, aimed specifically at supporting the sharing of qualitative and ethnographic data. The Data Story concept is motivated by three elements: 1. the increased attention of funding agencies and academic institutions on Research Data Management and Open Science; 2. our own work with colleagues applying ethnographic research methods; and 3. existing literature that has identified specific challenges in this context. Ongoing issues entailed in dealing with certain contextual factors that are inherent to qualitative research reveal the extent to which we still lack technical design solutions that can support meaningful curation and sharing. Data Story provides a singular way of addressing these issues by integrating traditional data curation approaches, where research data are treated as 'objects' to be curated and preserved according to specific standards, with a more contextual, culturally-nuanced and collaborative organizing layer that can be thought of as a "Story". The concept draws on existing literature on data curation, digital storytelling and Critical Data Studies (CDS). As a possible design solution for Research Data Management and data curation, Data Story offers: 1) a collaborative workflow for data curation; 2) a story-like format that can serve as an organizing principle; 3) a means of enhancing and naturalizing curation practices through storytelling. Data Story is currently being developed for deployment and evaluation.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**

Additional Key Words and Phrases: Data Storytelling, Data Reflexivity, Data Curation, Qualitative Data Sharing, Qualitative Research, Conceptual Design, Open Science, Research Collaboration

1 INTRODUCTION

For at least two decades, academic institutions have had to deal with the major changes implicated by a move towards the so-called Open Science agenda. This has the potential to reshape the cultural, organizational and infrastructural academic landscape [3]. In fact, most Western governments and all their major funding institutions fully embrace this agenda, with the clear intent of ensuring the verifiability of findings, promoting good scientific practice, and providing greater returns on public investment by encouraging data reuse [60]. To satisfy these objectives, data repositories and data centres are proliferating and many funding bodies now mandate the creation of research data management plans (RDMP) and the implementation of Open Data policies that embrace the "FAIR Data Principles" [62], i.e., research data deposited into archives should be Findable, Accessible, Interoperable and Reusable. Knowing how to efficiently structure, manage and curate data in order to fulfill expectations regarding long-term preservation, sharing and data reuse is becoming a sine qua non condition for receiving research funding. However, despite political and infrastructural efforts, the Open Science agenda remains some way from being realized and its ambitions have proven to be especially challenging for Humanities [24, 47] and Social Sciences (HSS) scholars [43] for whom these requirements are relatively new. Indeed, not all data are created equally and for some disciplines is much harder to adjust to these demands due to the nature of the data collected and the methods applied. Within the Social Sciences, researchers working with qualitative and ethnographic data are confronted with particular legal and ethical issues [19, 42], the personal character of the data can make researchers unwilling to share it in its totality, it can be hard to see what counts as metadata or how to curate qualitative data for sharing, and the sheer heterogeneity of data and data management practices can make standardization massively problematic [see e.g. 50]. As a result, the sharing and meaningful reuse of qualitative data remains rare, outside of teaching contexts [7, 8], nor are concrete solutions – beyond data archives or data repositories – being successfully implemented and regularly used as yet [37].

In our view, critiques of openness should be taken seriously. There is growing consensus that the mere release of data is not enough to realize the full potential of openness [42, 64]. In particular, open data portals or data archives are prone to becoming 'data dumps', where the number of published datasets is more significant than their quality or utility [see 45, for an account of the legal consequences]. Open data portals or data repositories are typically all about the structuring of data and the policies that surround it: how many datasets, how many formats, which open licenses and so on. While formats, standards and licenses are necessary for the long-term preservation of 'data objects' and their retrieval, there are still few design solutions that specifically support the practices and workflows necessary for interdisciplinary collaboration around those objects [23, 42]. In response to these challenges and critiques, we present an exploratory and conceptual design solution, called 'Data Story', that offers a particular way of curating and sharing heterogeneous data sources collected by ethnographically-driven research projects that can be seen to better resonate with the interests and expectations of qualitative researchers. The solution aims to support the partial curation of data by encouraging a pre-

selection of relevant data that researchers might wish to share that can then be contextualized by making use of storytelling practices. The concept grew out of a long-term engagement within an interdisciplinary Collaborative Research Centre, where we observed researchers working in interdisciplinary ethnographically-driven contexts and engaged in conversations with them about data their practices. The Data Story design can be seen as a way of building upon the current informal sharing practices we observed and of addressing the unsolved Research Data Management issues we surfaced.

The Data Story, as an exploratory and conceptual design solution, has its roots in literature relating to *data curation and sharing* [7, 8, 12, 55, 56]. However, it also takes inspiration from works relating to *data storytelling* [18, 34, 46], and *Critical Data Studies* [12, 13, 32]. Ethnographic and other qualitative data, historically associated with the social sciences but increasingly deployed in HCI and CSCW contexts, are inherently narrative in character. It follows that something akin to 'storytelling' might be an appropriate focus for the data sharing agenda. As we will be elaborating below, Data Story, as a concept, seeks to supplement traditional data curation approaches by adding a more contextual, cultural and collaborative *organizing layer:* "the Story".

2   RELATED WORK

There are three principal bodies of literature that delineate the research space this paper is addressed to. One of these reconstitutes data management as a sociotechnical issue and stands as a critique of approaches that assign a certain fixity to what counts as data. Another focuses more specifically upon the sharing of qualitative data and the unique challenges this can pose. The third is concerned with data narratives and data storytelling and the extent to which this has already featured in approaches to data management. We look at each of these in turn below.

2.1   Critical Data Studies and the myth of 'raw data'

As Dourish and Cruz [17] 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" [17, p.8]. We see this, above all, as an issue of rationale. *Why* is data collected, organized and represented in the way that it is? The desire to embed rationale into data can be traced back to the literature on 'design rationales' in the context of software design [11, 15, 35, 41]. As Lee [35] sums up the concern as follows: "Reuse/redesign/extension support... can serve as indices to past knowledge (similar designs, parts, problems encountered)." Big data, however, has prompted an epistemological shift away from relatively mechanical, model-based approaches to problems of storage and retrieval and towards a more practice-oriented view, at least for some. This has been a motivating force behind 'Critical Data Studies' [12,28] and various practice-oriented studies in CSCW, HCI and STS. 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" [33]. As noted, this was occasioned by a 'step change' in the production and employment of data.

At heart, critical approaches recognize that political, social, ethical, organizational, and economic elements shape data management as much as technical problems. If so, data can no longer be treated as having some kind of 'objective' status. Data, as Gitelman [26] has suggested, is always "cooked" and "raw data is an oxymoron". The construction and reconstruction of data formats depends on an array of factors, among others the cultural norms of the groups that

created them [op. cit.]. By way of example, Vertesi and Dourish [58] have shown how data management, including sharing practices, is mediated by the nature of research cultures. Thomer and Wickett [54] underscore this in their 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" [52, p.3]. As we shall see, curating for data work as a pluralistic and contextual endeavor has, as yet, not been fully realized.

2.2   Challenges for qualitative data sharing

Data sharing have been a topic of intense interest across a number of disciplines in recent years [4, 22, 27] motivated by the requests for Open Data increasingly mandated by all major funding institutions. Most of the literature points to the many unresolved challenges inherent in preparing data for sharing purposes. Documenting and providing sufficient context for others to understand how data has been gathered, analyzed and processed, and the lack of incentives and motivation on the part of the researchers are seen as the most critical issue [5, 63]. These apply equally to all types of data and disciplines. However, some disciplines such as the natural sciences have managed to better adjust to these new demands and with time have developed internal policies to ensure the sharing and eventually the reuse of research data [65]. For other disciplines these requirements are relatively new and researchers and institutions are still struggling to understand how to meet these expectations. In the Social Sciences specifically, it is recognized that the sharing of qualitative and ethnographic data presents particular challenges because of the epistemological, methodological and ethical complexities associated with this type of data that do not directly apply to quantitative data and/or other disciplines [43, 56].

The epistemological difficulty with qualitative data lies in the fact that it is challenging to grasp *what* the 'context' may be in any precise way and *how to* describe it [40]. Context determines whether something can be viewed as data or metadata and the "degree to which contexts and meanings can be represented influences its transferability" [10]. Others have questioned the legitimacy of data when removed from the original contexts, packaged in repositories, and disentangled from the knowledge and expertise of the researchers who performed the study [61].

With regard to methodological challenges, it is important to recognize the reflexive character of this type of research [14, 38]. The collection of qualitative data is inherently intersubjective, its analysis is iterative, and interpretation is always a key aspect of the work. Data are often rich with personal content and are neither collected nor analyzed in a linear manner [56]. Nor are many data collection activities targeted at sharing and archiving, so the resulting products are not well documented or formatted for others to use [30].

In relation to the ethical challenges, preserving the anonymity of study participants is of key concern. Informed consent stands to become significantly more complex if the sharing of the associated data for public consumption becomes commonplace [6, 44, 49]. As ethnographic approaches are generally based on a trust relationship between researchers and participants and can often focus on sensitive domains, there is a risk of this being undermined by the prospect of sharing data with unknown and potentially unaccountable parties. Anonymization of data (e.g. to comply with EU GDPR legislation) is typically offered as a solution, but the greater the amount of anonymization, the greater the risk of losing the contextual information needed to make sense of ethnographic data. There is also a lack clear standards regarding how to describe and prepare qualitative data for sharing [1, 56]. Data formats are difficult to identify due to the heterogeneous

nature and idiosyncrasy of researchers' data practices. Beyond this, time issues may also arise because "the burden of organizing qualitative data for inspection or reuse could easily exceed the work of writing the manuscript itself" [54, p.5].

Evidently, the sharing of qualitative data is anything but trivial. Effective sharing and potential reuse will remain problematic until there is an understandable and efficient way of preparing and curating data in a way that is aligned with researchers' practices and data work. In this way, we second the work of Rawson and Muñoz [47] who advocate for articulating new paradigm and practices in the field of Research Data Management that should support the humanistic way of dealing with data and its specific way of producing knowledge. As we shall see, Data Story offers an innovative and lightweight way of addressing some of these complex issues.

2.3 Data Storytelling: Guiding principles and insights

The social sciences and humanities have long stressed the role that narrative plays in human life, education and research. As Game and Metcalfe [25] 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'" [25, p. 65]. Social scientists tell these stories for a range of purposes. In doing so, they attempt to contextualize the 'data' that they work with. However, 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. Thus, 'context' 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 mechanism for reconstructing this process.

Digital storytelling simply refers to the digitally-mediated practices adopted by everyday professionals and organizations to tell a story. They can seek to stimulate emotional responses in recipients and can offer interactive elements. Digital storytelling can be found across numerous fields, including: therapy, education, arts and culture, library science, and management and business [2, 16, 30, 39, 48, 53, 57]. Over the last decade, the advent of big data and the data revolution [31] has led to western economies and governments becoming increasingly data-driven, leading to a growing focus specifically on '*Data Storytelling*' [46]. The main argument is that, to understand and use 'data' effectively, it needs to communicate a clear message (a narrative) in intelligibly human terms that enable us to make sense of it ('data sense-making') and understand why it looks (is reconstructed) the way it is.

At heart, data storytelling consists of three main elements: 1) explaining the context; 2) identifying a coherent narrative; and 3) providing effective visualization. Note, again, the emphasis upon *context* here, with it being the producer of the narrative who identifies the relevant context. At the same time, and as with all human communication, the narrative that is produced involves assumptions about its potential audience, which, in turn, recognizes the second active principle in data storytelling, i.e., *narration*. A narrative can stimulate learning, emotions, and drive action through discursive constructions. A story has a beginning and an end, it has a goal, sometimes a moral, and, as already mentioned, an audience that it is designed to engage. The power of narrative can help to share norms and values, develop trust and commitment, share tacit knowledge, facilitate unlearning, and generate emotional connections [52]. The third principle is related to *effective visuals*. Once data is analyzed, and the message and the story are developed, the data needs to be visualized accordingly.

Other literature, largely associated with critical data studies and STS, has focused on scientific storytelling and the creation of stories in a more qualitative fashion [29, 32, 36, 59]. Based on their fieldwork in the LTER network, Karasti et al. [29] showed how storytelling is integral to the practice of doing science, but also highlighted the challenges inherent in recalling, identifying and articulating stories while members are immersed in everyday work activities. Vertesi et al. [59] have also demonstrated that a narrative account of data management practices can help to uncover tensions in personal data management and allows the emergence of what they call "moral economy of data management", which express the "complexity, ongoing tradeoffs, emotional and reflexive components" of individuals' decisions and actions in respect to their own ecosystem of tools and data. Despite the recognition of narratives and storytelling as a useful mean to describe and talk about data and data practices little attention has been paid to how to translate such insights into design solutions. As Karasti et al. [29] point out, "Stories of everyday technical aspects of data work may be lacking due to data-work being considered something so mundane, even boring that it would be 'oddly inappropriate for an experienced worker to tell another experienced worker a story about daily routine [36]'. Yet it is just this tacit knowledge and contextual understanding that is essential for the analysis and design of 'narrative knowledge management systems' [29, p.30]".

To conclude, relatively little attention has been paid to data storytelling for design purposes and even less to its use in the construction and reconstruction of qualitative data for the purposes of data curation and sharing. We will be arguing here that the concept of a Data Story is particularly appropriate when making sense of qualitative research data. Storytelling can thus be used as an organizing principle when curating and sharing excerpts (snippets) of data from heterogenous data collections to facilitate its contextualization.

3 BACKGROUND AND APPROACH

As background to the Data Story, this section details the fieldwork and practical experience of working within a research infrastructure project (INF) that together guided its conceptualization and design. The project INF is connected to a Collaborative Research Center (CRC) that started in January 2016 and is still ongoing. The fieldwork was characterized by an ethnographic approach and comprised observations and semi-structured interviews, as well as long-term engagement and member participation. Interdisciplinary discussions concerning Research Data Management and data practices within CRC's projects took place regularly in the CRC and the principal author's involvement in these provided an opportunity for numerous formal and informal conversations with researchers. These conversations highlighted relevant RDM issues that make it difficult to meet the expectations of funding agencies for data sharing and reuse and therefore motivates the development of a new approach.

3.1 Empirical setting: The Collaborative Research Centre and the INF project

Our research took place in the CRC, an interdisciplinary research center consisting of 14 projects and more than 60 scientists from a variety of disciplines (i.e.: cultural studies, media studies, social sciences, digital humanities, engineering and computer science). Research projects in the center characteristically involve interdisciplinary cooperation, with most researchers using qualitative and ethnographic methods. This interdisciplinarity is further promoted by seminars, lecture series, workshops, PhD forums, and annual retreats, the latter being focused on discussing project updates and aligning research interests and findings. The CRC started in 2016 and completed its

first funding period in December 2019. A second phase began in January 2020 (funded until December 2023[1]). The funding agency, DFG (in English: German Research Foundation), first defined and adopted its "Principles for the Handling of Research Data" in 2010. These highlighted the importance of long-term archiving and the accessibility of research data, across all fields and disciplines. The principles are expected to be followed by all DFG-funded projects. A key element of our project, INF, has been investigating how to achieve this goal. INF's overall objective is to support the sustainable handling of research data, to develop and implement Research Data Management concepts, and to maintain the necessary infrastructure for the whole CRC. The INF project, which forms the principal background for this paper, therefore has a double focus: a) the provision of infrastructural services, led and represented by the IT service provider of the university; and b) design-oriented empirical research, conducted by the first author.

3.2 Fieldwork activities in the INF project and the research approach

Since 2016, the first author has engaged in monthly meetings with the IT service provider and its developers. These have included technical meetings to discuss unsolved RDM challenges and the brainstorming of design possibilities for a new research data infrastructure that could meet the expectations of the funding agency. One of the major challenges discussed was the lack of standard solutions for curating and sharing qualitative-ethnographic data. There was also a concern about how to resolve the top-down nature of the RDM mandate with the specific needs and interests of individual researchers. This prompted several rounds of empirical research to gauge how to proceed. Ethnographic observations and qualitative interviews were undertaken, largely between 2017 and 2019, that involved nineteen researchers representing all the major disciplines, roles and positions.

Table 1. List of the interviewees with their disciplinary background and academic position

ID	Pseudonym	Background	Academic Role
#1	Sophie	Media Science	Principle Investigator
#2	Joe	Media Science	PhD Student
#3	Alvin	Sociology	Post-Doc, Project Leader
#4	Lucy	Sociology	PhD Student
#5	Mary	Law	PhD Student
#6	Rupert	History	Principle Investigator
#7	Lukas	Sociology	Post-Doc, Project Leader
#8	Mark	Political Science	Project Leader
#9	Paul	Sociology	Principle Investigator
#10	Carl	Sociology	PhD Student
#11	Rob	Media Science	Principle Investigator
#12	Colin	History	Post-Doc, Project Leader
#13	Julian	Anthropology	PhD Student
#14	Aaron	Business Information System	PhD Student
#15	Philip	Computer science	Principle investigator
#16	Cliff	Business Information System	Post-Doc
#17	Susanne	Social Science	Principle Investigator
#18	Beth	Political science	PhD Student

---

[1] CRCs can be funded for up to twelve years across three separate evaluation stages (Phase 1; Phase 2 and Phase 3).

#19	Will	Anthropology	Principal Scientist

The fieldwork focused on understanding research data management practices from the bottom up, with a specific focus on documentation and sharing practices. More detail about these practices and the challenges we uncovered can be found in our previous work (reference omitted). The interviews revealed frictions between the expectations of the funding agency and researchers' actual practices. The funding agency's vision of RDM and the data life-cycle implied that research practices should be targeted at the long-term preservation of research data and ideally support both data sharing and reuse. In fact, while curation, sharing and consequent data reuse are central to the OS agenda, these practices are currently not much of a feature of qualitative research and are not well-supported by any of the tools qualitative researchers typically use. However, over the course of our long-term engagement, during which we undertook plenary discussions, group meetings, and supported researchers in drafting their Research Data Management plans, the researchers reported an interest in innovative solutions that might help them to represent and share their highly heterogenous research data in ways that would help them to organize it and underpin the work of collaborative interpretation. The Data Story concept was grounded in this apparent need. From it we came to see that the showcasing of data 'snippets' and the integration of storytelling practices could potentially support the organization, curation and eventually the sharing of research data, in greater synergies with the researchers' practices.

Fig. 1 – The first sketch of the Data Story idea, generated during a meeting in July 2019.
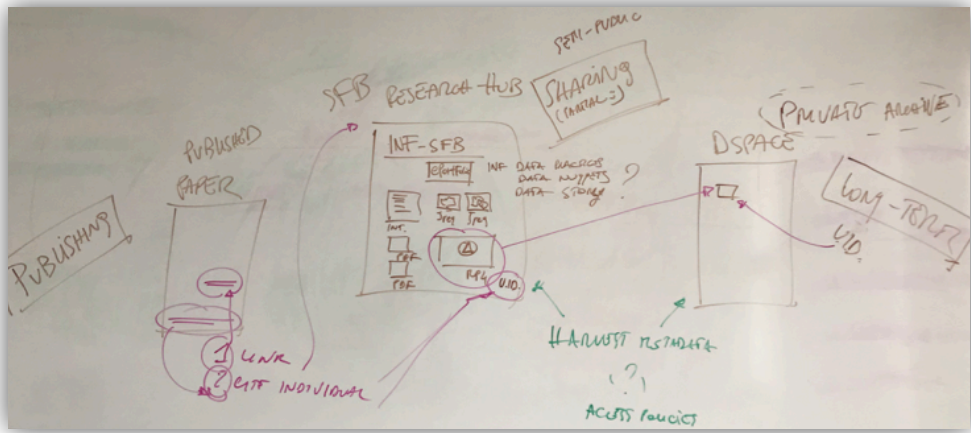
The first outline of the Data Story idea arose as a sketch[2] during a group meeting in July 2019 (Figure 1). It was inspired by the way that researchers were seen to share data snippets and engage with them on an ad hoc basis during internal meetings. This partial and purposeful sharing was the point of departure. We returned to it and developed the idea further by designing a low-fidelity prototype between January and March 2021. This integrated storytelling components as a way of providing contextual information and complementing it with basic metadata that could support data retrieval. Although the prototype has not been formally evaluated, its design is

---

[2] The middle side 'INF-SFB' represents the interactive interface of the collaborative platform 'Research-hub' through which share heterogenous data. 'Data nuggets' or 'data stories' are also imagined to be linked to published papers (left side) in order to make other researchers aware of these additional materials. The right side represents a long-term repository that could also be linked to a Data Story.

grounded in informal sharing practices and RDM issues reported by our CRC members during interviews and meetings. We will report on own our observations and related issues in the next section.

4 EMPIRICAL INSIGHTS

### 4.1 Data sharing: Informal practices and workarounds

Most CRC projects involve two or more disciplines working together, typically social scientists, anthropologists or media scholars working with computer scientists, designers and/or software developers. Methods for data collection are heterogeneous, often local to the disciplines involved. For ethnographically-oriented projects, the full data collection usually comprises interview files, ethnographic fieldnotes (often on paper), archival documents and other types of media, such as audio, pictures and videos. Data is often stored on personal hard drives and/or in Cloud Systems (like Dropbox, OneDrive, Sciebo or Sharepoint). Generally, only the researcher(s) who actually engage firsthand in the fieldwork activities have full access to it. Different data types are distributed across several repositories and almost no one in the project has an overview, not least because few people anticipate a need for full access. Instead, during research meetings, data that exemplifies putatively 'important' themes is usually presented. This data is pre-extracted from larger datasets with a view to meeting presumed analytic agendas, engaging in collaborative interpretation, discussing major findings, developing design ideas, structuring publication outcomes, and so on. Thus, it is common practice to share 'data snippets' in collaborative analysis sessions with members of the same project (but with different disciplinary backgrounds) and/or with researchers from other projects. These snippets of anonymized data are often enriched with contextual information (e.g.: time and place of collection, atmosphere, informant background, etc.) and sent to participants via email a few days before the analysis session.

At the very beginning of the actual session, a *narration* or, if you will, a *story* that contextualizes the data is often provided by the data collector in written form (i.e. as text), and/or in oral form. The data itself is then often displayed to guide the conversation and promote interpretative work. Through these oral and written narratives, qualitative data is constantly evolving and being co-constructed in a collaborative effort that can engage team members, research advisors, student assistants, fellow researchers and even study participants. Claude, a PhD student told us: "*I like storytelling and I even catch myself sharing data that way, I share snippets of my fieldwork and I add some sort of storytelling to it to give others an idea of what I did or what's the background to a short piece of data I might want to talk about*"*(Claude, PhD student from HCI, forum discussion on May 2020).* Full ethnographic datasets and, in particular, fieldnotes, are not fully shared, due to concerns about both its potentially sensitive nature and time constraints. As another PhD student put it: "*It doesn't necessarily help if I make my whole notes accessible to all the team members, it will take too much time to read it all, and also, I wouldn't want that either, because it's a rather personal thing*" *(Julian, PhD student from anthropology, interview on April 2019).* So, even among colleagues from the same research team, qualitative data is often shared only partially. There is pre-work involved in selecting the most relevant data. This may then circulate via email, but also often ends up on different commercial software platforms, like Dropbox or Google Drive, where researchers organize it to foreground what they consider to be most important. Many researchers feel this informal data sharing practice is 'not ideal', mainly because it implies the use of commercial platforms, but also because it results in different chunks of data being spread across multiple platforms or file sharing systems without any consistent

structure. It should be emphasized that researchers resort to using these commercial platforms because they are the only obvious solutions that can efficiently support simultaneous collaborative interpretative work around written data narratives.

As platforms change and evolve, so researchers have to constantly come up with new techniques and tools to assist them in communicating ideas and interpreting data. As an example, an historian made use of a Trello board (see Figure 2) to collect and structure heterogeneous data sources. The most important pieces of files, pictures and historical documents were organized into thematic sections, annotated and collaboratively discussed with student assistants.

Fig. 2 – Picture taken by the first author during an interview. It represents the Trello board of a historian who used it to organize heterogeneous data collaboratively collected together with his team.
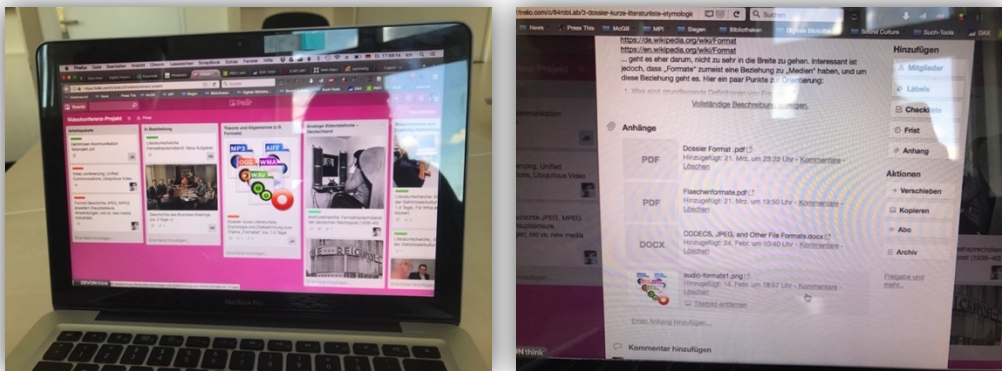
Here, Trello provided a great workaround for structuring heterogenous data sources. However, each data snippet or document is not made searchable, single data entries cannot be easily exported from the tool in order to continue analytic work, nor can the data be officially shared, presented or cited by another scholar. This specific researcher was in fact interested in experimental publication formats that would allow him to share and publish oral history interviews as video material together with transcripts and other supporting data.

Something important we found was that qualitative researchers are not opposed to the notion of data sharing in principal. Rather, they actively want to learn from one another and seek to understand what type of data other projects collect and how to organize, share and represent research data in innovative ways: *"you can also suggest (...) to talk to other projects who have similar research data in order to maybe, yeah, think about standardization. Do we need that, do we not need it because we're so small, are there even standards for archiving these types of research data? (...) also, for presenting this invisible work, because making interviews is very time consuming, but it doesn't really show a lot, so to have something like a representation of that would be great"* (Colin, Postdoc from Media History, Research Data Management plan meeting on January 2020).

Based on these observations, Data Story started to emerge as a solution that researchers could engage with at any stage of a project in order to represent snippets of heterogenous research data, engage in discussion concerning data interpretations and develop bottom-up curation standards. What we particularly took from the above is the fact that there are already mechanisms that qualitative researchers have in place for sharing data. However, what does not happen is the sharing of all of the data all at once, if ever. Rather facets of their data are shared, having been pre-treated in certain ways, and those facets are embedded in narrative structures that *premise*

the data in certain ways, according to its expected recipients, just as stories are shaped for their anticipated audience in certain ways. Another key observation was that qualitative researchers currently struggle to find consistent ways of doing this, but rather adopt formats, structures and platforms in a piecemeal fashion according to whatever currently seems to be at-hand. Data Story, as a concept, focuses on this nodal point, between the data and those with whom it might be shared. It is not an end-to-end solution, but it seeks to draw upon what already happens naturally and to imbue it with more structure and to make it more conducive to meeting some of the more formal demands associated with the Research Data Management agenda. Thus, it might be seen as a way of facilitating the readier sharing of qualitative research data than is currently the case. In the next section, we explore how Research Data Management, as it stands, is seen primarily as a source of tension.

4.1 Research Data Management issues

When the CRC started its "first phase" in January 2016, not all projects were fully aware of the DFG agenda relating to the long-term preservation and accessibility of data that would imply commitment on their part to share data. CRC members showed skepticism regarding this agenda and questioned the expectations. As Carl, a sociologist, told us:

"now we did the interviews and we didn't even know if there was going to be one repository or what that would look like (…) yes, we are a bit skeptical because again when we filed the application for the research project nobody came up with the idea yeah somebody would eventually need to anonymize all that data. Eventually you need to have somebody who does that and that work power was not sort of calculated within the original calculation right?!" (Carl, PhD student from Sociology, Interview on July 2017)

At the time, the university had no Open Access policy guidelines[3] and no long-term repository to offer as an archive service to CRC projects. The INF projects wrote the policy in late 2017 and the repository infrastructure was finalized in June 2021. Meanwhile researchers were assisted with the creation of RDM plans for phase II, but no research projects formally agreed to allow their data to be publicly accessible. They only committed to engage with the long-term archive and even the archival process generated a number of concerns, especially with regard to metadata and the documentation to be deposited with the data. Some researchers even said they would not provide any documentation because *"that is a practice currently not in place".* Some researchers were curious to know more and wanted to learn how to create the right metadata and documentation, but they were disappointed by the replies they were given. The IT service provider, for instance, suggested they use the Dublin Core metadata and simply shared a link with them. After consulting the link, they came back to us and said: *"we literally have no idea how and when we should be using this?! The standard names and definition are expressed in a very technical language that makes it difficult to understand what is asked exactly, what is the coverage?"* (Lukas, Postdoc from Sociology, RDM plan meeting on December 2019). The proposed metadata had no clear link with the data the researchers collected in folders via file sharing repositories or in their personal hard drives. This made it impossible to support a workflow where documentation practices and metadata entries could be embedded in the everyday business of their data work. The solutions offered at the time largely focused on long-term preservation and sharing, not on the more immediate problem of how to choose what data to share and how to share it as a part of one's everyday work.

---

[3] These were finally implemented in March 2017

A particularly evident problem was the heterogeneity of people's data and the metadata used to organize it. As one sociologist put it, *"these are some protocols of the interviews with some information, like the name, the age, what the people are doing, how the interview came about, what the communication was before the interview, what the interview was like, where it took place, how the atmosphere felt, were there breaks or pauses for various reasons, what the people looked like, how I felt, how they seemed to feel, and so on ... if we share data it needs to have this information, along with things like the questions we asked."* (Alvin, Postdoc from Sociology, Interview on April 2017). Other data often found in qualitative work includes descriptions (written and pictorial) of physical layouts, the positionality of the researcher, and difficulties encountered. Some researchers even incorporated information about what they had failed to find out. We would argue that contextual information of this kind is not easily represented in existing metadata structures. Some approaches, such as ethnography, are not at all commensurate with the step-by-step process idealized in the data life cycle. For example, initial analysis and interpretation of 'data' often begins as the fieldwork itself starts and continues until publication. Interpretation, reflection, and documentation also continue throughout the research process, incrementally adding descriptions to the materials collected, which are often enriched with personal reflections and emotions. Sharing, then, is difficult without some level of 'curation' and pre-selection of data. Currently, no processes exist to afford ongoing curation and partial sharing in this way, even though this is a routine feature of qualitative research. The expressed need for flexible contextual metadata, pre-selection and partial sharing all resonates strongly with the existing practices and requirements identified in Section 4.1.

Some researchers wanted to explore solutions that would allow them to record aspects of the analytic process in support of methodological reflexivity: *"that's something I am super interested in. How do you kind of make sense of the different data sources that you are working with? (...) How do you make sense with it in a research process, what kind of decisions are being made and where? And so being kind of reflexive and accountable of your methodological steps is something that I am interested like both like intellectually and also then that motivates to open up not only the data but also the decision process that comes with it"* (Sophie, PI in Media Science, Interview on April 2017). Some researchers were clearly interested in having a tool that would allow them to organize different data sources and support analytic reflection. They argued that this would pay dividends by making sharing and courting feedback more straightforward. Data Story was therefore also focused upon providing such tools to support both ongoing and completed research. In particular, we envisaged an interactive interface that could present their data for comparative purposes, allow for intermediate feedback, promote the ongoing evolution of research data, and potentially provide an alternative publication format.

5 THE 'DATA STORY' CONCEPT AND ITS DESIGN

The preceding materials provide a backdrop to the development of the Data Story concept and its design. This concept takes the notion of a 'story' as a design metaphor and uses it as a source of inspiration for the representation, organization and description of *partial and situated* research data to be shared with colleagues, and/or with external audiences. As mentioned above, it is not an end-to-end solution. Instead, it takes existing practices, concerns and requirements as a point of departure and seeks to facilitate the establishment of curation and sharing practices. The core idea is to showcase anonymized 'data snippets' (interview excerpts, pictures, videos, sketches or any other relevant material) that are organized in such a way as to elicit storytelling practices (in oral and written form) to contextualize the data. Above we noted that we recurrently observed

researchers telling stories about their data, but in a relatively unstructured way. Our design seeks to give more structure to that practice, while affording other aspects of the data curation process that meet researchers recorded wishes.

The concept draws on all three affordances of data storytelling identified in the literature by providing: a) a way to contextualize collected data; b) a narrative structure to demonstrate its analytical potential, c) a vehicle for the integration of additional representational elements. We discuss below how 'Data Story' was envisaged in accordance with these principles. It should be emphasized that Data Story has not yet been deployed and evaluated. To date, we have only developed a low-fidelity prototype[4]. Therefore, at present, it only has the status of being a conceptual design, albeit grounded in our empirical work. We plan to implement this design in the up-and-coming months as an *independent module* in an existing and established platform called 'Research-hub', which is built for team collaboration and sharing and that is already used by multiple research groups in our university.

5.1 Research-hub

'Research-hub' is a customized platform based on Humhub open source software for team communication and collaboration (see https://www.humhub.com/en). Research-hub is already in use in our university as a resource for research project management, academic collaboration (collaborative paper writing, reading groups, etc.) and teaching. In the future, the goal is to also support curation and data sharing practices as well. The platform has a three-level hierarchy: 1) *User profile*; 2) *Spaces*: smaller collaborative units (e.g., research projects); and 3) *Communities*: larger organizational and institutional units (e.g., Departments, or Research Centres). Spaces are linked to a specific community if they belong to the same institution. We intend to develop Data Story as a module that will be connected to the User and Space levels (i.e., smaller collaborative units). Certain outputs - once published - will also be displayed at a Community level for broader sharing within the institution. Hanging Data Story off of Research-hub facilitates easy cross-discipline, cross-project and cross-department sharing. This reflects the existing character of many meetings within which data-sharing takes place. It also stands as an example of what the Open Data agenda might be seen to be about, but in miniature.

5.2 Data Story Design in a nutshell

Data Story provides a preliminary structure or template to help researchers organize and describe the context of a specific study by making use of written narratives (stories). The interface is organized into chapters, so that shared data can be sorted into sections, aiding navigation through the story. The sequential organization of the chapters creates a timeline of actions, events, and decisions regarding the study being shared. Each chapter might contain multiple documents and 'data snippets' that help to clarify the overall story. Questions and tips are highlighted in the interface of each chapter to support researchers in crafting their own narratives, to encourage reflexive thinking and elicit discussions. Story authors can add few selected metadata to the 'data snippets' to enrich the explanation (and support future retrieval), but in a way that allows for learning about and questioning the role of metadata as well. Authors can also introduce themselves and their research institution and give their contact information, etc. This is needed

---

[4] **The full prototype can be accessed at this link**: https://www.figma.com/proto/TtFgWU2Oau7njVk9klyZgI/Data-Story-Module?node-id=209%3A12&scaling=scale-down&page-id=0%3A1&hide-ui=1

to connect a Data Story with a specific researcher or research team (so as to be publicly acknowledged and possibly contacted).

Fig. 3 – Example of a Data Story structure: chapters' overview and related metadata.
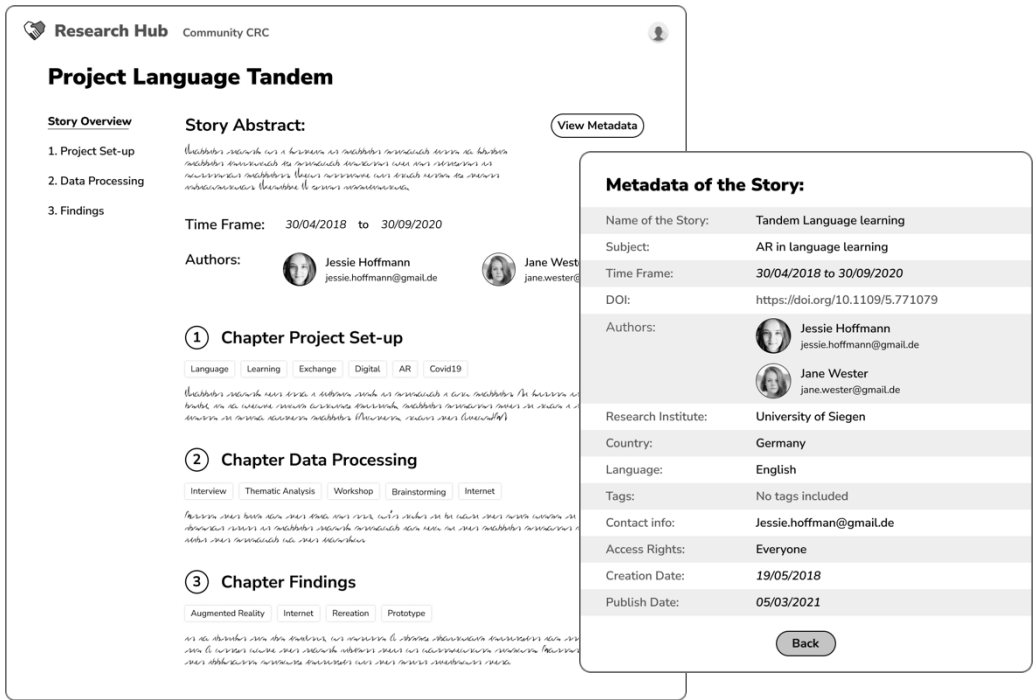
To exemplify the possibilities, we provide a possible structure for three different chapters within a Data Story (see Figure 3): (1) project set-up; (2) data processing (with snippets of anonymized data); (3) main findings. Each chapter provides a focused insight into the study conducted and suggests a narrative structure threaded through the chapters. There is some general information regarding the story that is provided in the overview screen. This can give information about the time frame and the project to which it belongs (a single publication, a complete research project, a PhD dissertation, etc.). Across the chapters, authors are encouraged to enrich the Data Story with various kinds of contextual information that echoes many of the practices we observed during our own fieldwork, where not just pre-selected data was made available to prospective meeting attendees, but also information about it, such as when and where it was collected, the informants, etc. This can serve to overcome a number of the issues we identified, such as the fact that data snippets are not currently provided in ways that make them searchable, easy to export to continue analytic work, or open to broader sharing or re-use. It also oversteps many of the current concerns being expressed about metadata and documentation by providing a natural way for this to be embedded in the preparation of data for sharing that respects its potential heterogeneity.

5.2.1 The project set-up chapter. The project set-up chapter introduces the overall story outline, thus providing an understandable context for the study. Information related to the study's domain, topic, research questions, methods, author contact information, motivation and aims can all be included. Tips and questions are highlighted in the interface. In this chapter, researchers

who write the story are encouraged to consider the following questions (and include their answers in their story narrative):

- How does your story start and where is it situated?
- What is the topic of the story?
- What is/are the research question(s)?
- Why are you sharing the story? What is the goal?
- For whom would this story/study be interesting?
- Who would be interested in your data?

5.2.2 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 analysis process. As with the project set-up chapter, the processing narrative is aimed at resolving common queries to support analytical reflections of the shared data nuggets. This feature of Data Story took inspiration from, and is actively designed to reflect existing practices.

We saw how data was pre-selected from larger datasets to illustrate putatively important themes, in a way that could support collaborative interpretation, discussion, and the development of further ideas. It also avoids a need for researchers to proceed in radically different ways because they are already investing effort in the pre-selection of the most relevant data to support their interactions with other researchers. One of the key advantages to proceeding in this way is that it may serve to eliminate the current tendency for data snippets to be spread across disparate platforms and file-sharing systems without any consistency of structure.

In this chapter, sub-sections can be created to categorize and group data, based on the data type and methodology, thereby easing navigation. Authors are advised to create and fill the sub-sections with relevant data in a way that supports the storyline and its sequence, with sub-sections being ordered sequentially (see Figure 4). Authors can position and relocate sub-sections by simply dragging them to their desired location on the storyline. Example of data types in this sub-section are: informed consent; interview guidelines; observations; interview data; focus groups results; workshop protocols; evaluation outcomes; etc. Customized sub-sections can be created where desired data categories are missing.

Data Story supports the sharing of different data formats. Some snippets might be extracted from a text file and 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 materials, etc. All of this can be seen to reflect a need to support existing heterogeneous data collection practices with a variety of data formats. As before, the authors are provided with a list of questions to help them structure the story, support the 'sense-making' of the shared data, and enrich the contextual layer. For example:

- What methods were used to collect and analyze the material?
- When was the data collected (timeframe)?
- What data types have you considered during the analysis?
- Keep in mind: what/why/with whom are you sharing?

Only selected and anonymized data will be displayed. There are three principal reasons for this, each clearly articulated in the empirical insights presented above: 1) to protect study participants and avoid the disclosure of any private and sensitive information; 2) to decrease 'data overload' by encouraging researchers to display only the most relevant data; 3) time constraints - it is not possible to provide a complete and carefully crafted narrative in a relatively short period of time that will adequately contextualize all of the data collected during a study. It also in no way breaches existing practices where the original data is stored on personal hard drives or in the cloud and is only accessible to the original data gatherers.

Fig. 4 - Data processing chapter and its structure once completed. Each sub-chapter contains relevant data and contextual information that helps to make sense of the data itself and of the methods applied.
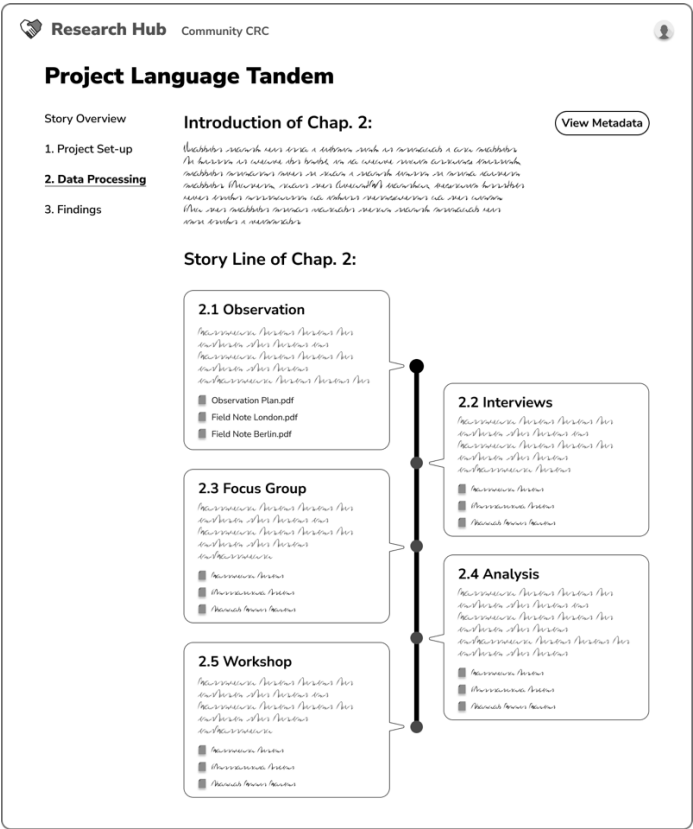
5.2.3 The findings chapter. Last but not least is the findings chapter, where the narrative ends and future plans are explained. Published materials and citation and review data can be included in this chapter. Again, guiding questions and tips are visible upfront to help researchers structure the information and narrative, e.g.:

- What findings came out of your data?
- To whom/in which fields is this data story specifically useful?
- Bring the story to an end.

5.3 Integrating metadata standards

In general, Data Story gives the option to annotate, tag and add metadata to every chapter. Keywords and relevant tags can be assigned to both the story in general and to individual chapters. This provides a quick overview of the general context and topic of the story. Data Story suggests few basic metadata (i.e.: the Dublin Core or DDI) as a standard source for elements. They can, however, be edited quickly and/or a new folksonomy can be created to explain the data. As mentioned earlier, Data Story invests effort in bringing the data and its metadata together by integrating many of the important metadata fields in its interface. This makes metadata an important pillar of the narrative and a driver of discussions. It promotes 'data literacy' and

'awareness' by providing an opportunity for researchers to learn about and reflect on the role of metadata and finally adapt it to their needs. We also envision that the metadata elements will change depending on the data type: i.e., some will be suggested for interview snippets, but different ones for ethnographic notes, focus groups, design sketches, etc. Further research is needed to identify just which elements best match different data types.

5.4 Supporting processual workflows: plugin solution

Data Stories can be posted with key data and story milestones at any time throughout the course of a study. In fact, Data Story aims to promote ongoing curation activities as a feature of everyday workflows. To achieve this, Data Story will be connected to routinely-used tools for collecting, analyzing and processing data. We therefore envision a plugin solution. The plugin can be connected to text-editing software like Microsoft Word, data analysis tools like MaxQDA, literature management tools like Citavi, cloud storage tools like Sciebo[5], etc. The idea is to provide researchers with an opportunity to feed their Data Stories with new input at all times by creating direct connections between Research-hub and their own data stores. In this way, researchers can select key segments (text, files, etc.) while organizing and analyzing their data and send them to a Data Story as '*data snippets*'. They will also be able 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 in Research-hub. They will be located in the data processing chapter unless directed otherwise.

In this ways, Data Stories can be assembled piece by piece as a natural extension of researcher's own data processing workflows, instead of trying to organize the data at the end of a study. This has a better fit with existing approaches where it is important to allow for interpretation, reflection and documentation to continually be enriched throughout the research process.

5.5 Publishing: DOI and accessibility rights

Once researchers have completed a Data Story and feel secure with the provided data and narrative, they will be able to publish it. Once published, the story will be visualized in the respective community. A DOI (Digital Object Identifier) can also be automatically assigned to the Data Story. A Data Story's DOI can also be promoted in papers, so that potential collaborators or interested parties can see additional data. Individual sharable links will also be automatically generated for single data entries so that researchers can give others direct access to a specific data snippet. It is up to authors to decide the amount of data to include in a Data Story. Some authors may decide to share very small snippets, others more substantial chunks. Critically, they can also decide what to share with whom. They can share certain parts with some recipients and other parts with some other audience, all within the same Data Story. This is facilitated by having different accessibility rights provided in the Data Story for each data snippet added to the storyline.

Data Story will be accessible within the existing Research-hub platform via a web browser. By integrating it into this platform, it will be possible to engage in discussion (if desired) with people interested in the data. All in all, Data Story hopes to trigger collaborative discussion, negotiation, awareness, sense-making and reflexivity around shared data by whichever parties have an interest, thus neatly tying it into the Open Science agenda.

---

[5] Info on Sciebo: https://hochschulcloud.nrw/en/index.html

6 DISCUSSION

Above, we have described an approach, inspired by storytelling insights and designed to support a collaborative workflow for the curation and sharing of data which can be used in conjunction with more standard approaches and data descriptions. We have shown how we might, in this way, address some of the more problematic aspects of Research Data Management and how to meet the expectations of the Open Science agenda. We discuss below some of the key points that can be seen to arise out of adopting this approach. First, we consider how Data Story can serve to articulate a RDM-related collaborative workflow. We then look at the specific advantages of promoting the notion of a 'story' as an organizing principle. We conclude by reflecting upon how Data Story, while not necessarily reducing the overhead of data management, can play an important role in naturalizing the RDM process.

6.1 Data Curation as collaborative workflow

In our fieldwork, we have noted how the curation and sharing practices implicit in the Open Science agenda are currently not yet visible or are only being performed in very haphazard ways. At the same time, however, we have seen that there is actually a willingness to share data amongst qualitative researchers we engaged with and, in particular, an interest in how to undertake collaborative work around their data. Data Story actively evolved out of our observations as a potential solution to manage evident RDM issues, to do that in ways that resonate with existing concerns, interests and practices, but, at the same time, to bring qualitative data curation closer to what is increasingly being demanded by funding bodies. While Data Story is not intended to offer a comprehensive RDM solution, it does view RDM and data curation as a process to be embedded in daily practices within a collaborative workflow. In fact, current tools are not yet interconnected in a workflow and miss to offer the opportunity to engage with curation elements such as adding metadata and annotations 'on the go'. Data Story integrates those missing elements but also provides a 'mechanism for narration' with an interactive interface that helps researchers to contextualize and organize their own data with written narratives, a practice more aligned with their way of doing research and dealing with data. Composing Data Stories will still be time consuming for researchers and curation practices will remain an overhead, however, we hope researchers will have the opportunity to gain personal benefits such as organizing heterogeneous data but also structuring relevant findings and analytical reflections to be used in collaborative discussions, publications and future work.

As Birnholtz and Bietz (2003) have already suggested, to get more effective data sharing systems, designers and IT developers need to go beyond current metadata models and take into account social interactions around data abstractions. Data Story takes this recommendation seriously and stresses the collaborative and social dimensions inherent in data and data practices. At this stage, we cannot yet anticipate how the narrative work will actually play out in practices, which difficulties researchers will encounter in engaging with its workflow and with whom Data Stories will actually be shared. However, we believe that a system like this would be a first attempt to bring the invisible data work to the forefront, to embrace the difficulty of making curation activities 'fit' the realities of local practices and the contingent nature of sharing practices. In this sense, Data Story promotes data awareness and reflexivity, and involves making curation activities and their concerns, technicalities, and specificities visible, while articulating workflows and processes that encourage social interaction and collaboration around data. An important part of how this will be accomplished is the embedding of Data Story within an existing platform, 'Research-hub', that is already purposed for communication, collaboration and sharing across a

diverse set of research groups and disciplinary interests. Data curation and sharing practices are, in a sense, a new concern for some academic groups. As such, the concern has yet to be consolidated and supported by collaborative tools. As Mosconi at al. [42] have already noted "the institutionalization of data curation practices [...] requires a better understanding of the use of data in practice but also the development of reliable infrastructure and tools built in a way to help negotiate OS objectives, stimulate self-reflective and learning processes and support discipline-specific data practices" [42, p.781]. Data Story is an attempt to address all these issues within the CSCW research tradition and its core themes [9].

6.2 The story as organizing principle

The practice of storytelling invites data handlers to think about their data in a way that encourages data reflexivity. Reflexivity has a special status in HSS disciplines, where there is a particular focus upon the relationship between researchers and their data. Through the organization of data snippets as data stories, researchers are specifically invited to reflect upon: 1) what they are sharing (i.e.: what are data, what are metadata etc., what methods were applied, etc.); 2) who they are sharing with; 3) why they are sharing in the first place; and 4) how the data's recipients might understand it. As our fieldwork highlighted, reflexivity is not typically prompted by standard approaches to data curation due the technical and generic language in which metadata standards are normally expressed. Therefore, with Data Story we wish to support data sharing practices while at the same time encouraging greater reflexivity during the process of curation and sharing.

We envision how Data Stories can be seen as a potential solution to the challenges outlined in section 4.2 by accompanying the self-archiving process. Writing a Data Story could be seen as a first step prior to depositing the data into an official archive. The storytelling approach is not intended to replace data curation activities or data curators. Instead, Data Story can enhance the curation process by opening up the 'black box' of research and providing the cultural and contextual circumstances in which data are generated, enabling this to happen during the research process, before formally archiving the data. It can fit organically between the moment of data production and its formal archiving, acting as an interface that can meet the practical issues researchers are currently confronting. As indicated in the empirical insights (section 4.1), researchers already tell stories about their data in their everyday work and undertake some of the activities encompassed within Data Story. Data Story takes those existing practices and concretizes their formulation and pursuit within a visible narrative. It also facilitates the sharing of those everyday stories about data across larger cohorts of researchers. In research about situated storytelling practices [51], much is made of how storytelling provides a mechanism for the sharing of experience. There is a sense in which Data Story builds upon that sentiment by recognizing the power stories have to promote sharing and engagement. The same research also emphasizes how storytelling practices require mundane competences that most people engage in willingly. It doesn't prevent data management and curation from being work, but it does make that work more familiar and routine.

Longer term, our approach will be suitable for data reuse. The central question in this respect is, how does Data Story provide a narrative that not only contextualizes the production of data but also renders it relevant for those who might use it. There is no simple answer to this question, for the value of data in reuse depends as much on the reasons for reuse as it does on the reasons for its production. Nevertheless, Data Story can do a number of useful things as its chapters' structure affords certain data relevancies: the project set-up can tell re-users why the data exists

in the first place, its potential value in relation to existing knowledge, and information about the disciplinary origins of researchers; the data processing and the 'snippets' can answer some of the queries re-users may have about the methods adopted, the amount of data and its format and give examples of the data, etc.; the findings can provide a link between snippets and results, enable judgements about accuracy, reliability and validity to be made, reveal literature deemed to be relevant, point to reviews of the work, suggest options for future progress. Finally, the overall narrative positions both data creators and potential re-users as active agents in the construction of meaningful data. Stories invite both data creators and the data 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.

7 CONCLUSION AND FUTURE WORK

To conclude, organizing, communicating and understanding data are crucial issues in a 'datafied society' [21]. Yet, in our digital world it is not always clear what counts as data, how best to make sense of it, and what is at stake when it is put to use [32].

Data Story aims to foster exchange around data storytelling that can serve as medium to explore data sense making, support data awareness and reflexivity. Although we have focused here upon qualitative data, the concept is, in principle, agnostic as to what is deemed to count as data. Instead, it is able to embrace a plurality of data practices and approaches. Data Story drew upon insights from CSCW, Critical Data Studies and related disciplines. These emphasize the sociality of work practices and the co-construction of meaning. Through Data Story, we want to promote more inclusive data practices that embrace a broader audience and provide diverse and faceted entry points for personal explorations. Our wish is to promote a smooth transition toward open science principles while remaining "as open as possible and as closed as necessary" [20]. At present, Data Story remains a conceptual contribution - it has not yet been deployed or evaluated. However, we plan to develop the work further by implementing and evaluating it across a range of projects. In this way, we hope to refine the concept and to gain deeper insights into how it might best support researchers and the recipients of their data. Clearly, we cannot wholly predict what the outcomes of this process might be, though we can speculate. For instance, innovations of this kind might form the basis of new publication formats in the longer term and help to incentivize the work of data curation, which is currently largely seen as unrewarding.

As Rob Kitchin has pointed out, the cooking of data does not take place in a vacuum. Data-driven endeavors are socio-technical in nature. They are as much a result of human values, desires and social relations as they are of scientific principles and technologies [32]. Such a view, we would argue, is fundamental to the CSCW tradition and we would encourage researchers in the field to use the Data Story concept we have presented here as a starting point for examining how alternative approaches to data sharing and reuse can be developed. How data is socially constructed, and how the stories researchers naturally tell about their data feed into its subsequent sharing, reuse and appropriation, should be a fertile field for CSCW research and may have much to offer in turn about effective data design.

REFERENCES

[1] Alison L. Antes, Heidi A. Walsh, Michelle Strait, Cynthia R. Hudson-Vitale, and James M. DuBois. 2018. Examining Data Repository Guidelines for Qualitative Data Sharing. *J. Empir. Res. Hum. Res. Ethics* 13, 1 (December 2018), 61–73. DOI:https://doi.org/10.1177/1556264617744121

[2] Helen Barrett. 2006. Researching and Evaluating Digital Storytelling as a Deep Learning Tool. *Soc. Inf. Technol. Teach. Educ. Int. Conf.* 2006, 1 (2006), 647–654. Retrieved August 31, 2021 from http://www.editlib.org/p/22293/

[3] Sönke Bartling and Sascha Friesike. 2014. *Sönke Bartling & Sascha Friesike.* Springer Nature, London.

[4] Harry van den Berg. 2008. Reanalyzing qualitative interviews from different angles: The risk of decontextualization and other problems of sharing qualitative data. *Hist. Soc. Res.* 6, 1 (January 2008), 179–192. DOI:https://doi.org/10.12759/HSR.33.2008.3.179-192

[5] Jeremy P. Birnholtz and Matthew J. Bietz. 2003. Data at Work: Supporting Sharing in Science and Engineering. *Proc. Siggr. Conf. Support. Gr. Work* (2003), 339–348. DOI:https://doi.org/10.1145/958160.958215

[6] Libby Bishop. 2009. Ethical Sharing and Reuse of Qualitative Data. *Aust. J. Soc. Issues* 44, 3 (March 2009), 255–272. DOI:https://doi.org/10.1002/J.1839-4655.2009.TB00145.X

[7] Libby Bishop. 2012. Using archived qualitative data for teaching: practical and ethical considerations. *Int. J. Soc. Res. Methodol.* 15, 4 (July 2012), 341–350. DOI:https://doi.org/10.1080/13645579.2012.688335

[8] Libby Bishop. 2014. Re-using Qualitative Data: A Little Evidence, On-going Issues and Modest Reflections. *Stud. Socjol.* 3, 214 (2014), 167–176. Retrieved August 31, 2021 from http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-657a6a2d-6222-4613-a6f3-85e17b08f124

[9] Jeanette Blomberg and Helena Karasti. 2013. Reflections on 25 Years of Ethnography in CSCW. *Comput. Support. Coop. Work* 22, 4–6 (August 2013), 373–423. DOI:https://doi.org/10.1007/S10606-012-9183-1

[10] Christine L. Borgman, Andrea Scharnhorst, and Milena S. Golshan. 2019. Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *J. Assoc. Inf. Sci. Technol.* 70, 8 (August 2019), 888–904. DOI:https://doi.org/10.1002/ASI.24172

[11] Janet E. Burge, John M. Carroll, Raymond McCall, and Ivan Mistrik. 2008. *Rationale-based software engineering.* Springer Science & Business Media. DOI:https://doi.org/10.1007/978-3-540-77583-6

[12] Craig M. Dalton and Jim Thatcher. 2014. What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data.' *Soc. Sp.* 29, (2014). Retrieved August 31, 2021 from https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care

[13] Craig M Dalton, Linnet Taylor, and Jim Thatcher (alphabetical). 2016. Critical Data Studies: A dialog on data and space: *Big Data Soc.* 3, 1 (June 2016). DOI:https://doi.org/10.1177/2053951716648346

[14] Charlotte Aull. Davies. 2008. *Reflexive ethnography : a guide to researching selves and others.* Routledge.

[15] Peter Demian and Renate Fruchter. 2009. Effective visualisation of design versions: Visual storytelling for design reuse. *Res. Eng. Des.* 19, 4 (2009), 193–204. DOI:https://doi.org/10.1007/S00163-008-0051-4

[16] Stephen Denning. 2006. Effective storytelling: Strategic business narrative techniques. *Strateg. Leadersh.* 34, 1 (2006), 42–48. DOI:https://doi.org/10.1108/10878570610637885

[17] Paul Dourish and Edgar Gómez Cruz. 2018. Datafication and data fiction: Narrating data and narrating with data: *Big Data Soc.* 5, 2 (July 2018). DOI:https://doi.org/10.1177/2053951718784083

[18] Nancy Duarte. 2019. *Data story : explain data and inspire action through story.* Ideapress Publishing.

[19] Igor Eberhard and Wolfgang Kraus. 2018. Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien. *Mitteilungen der Vereinigung Österreichischer Bibl. und Bibl.* 71, 1 (2018), 41–52.

[20] EC - European Commission. 2016. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020.*

[21] Karin Van Es and Mirko T Schäfer. 2017. *The datafied society. Studying culture through data.* Amsterdam University Press.

[22] Ixchel M. Faniel and Trond E. Jacobsen. 2010. Reusing Scientific Data. In *Computer Supported Cooperative Work,* Kluwer Academic Publishers PUB879 Norwell, MA, USA, 355–375. DOI:https://doi.org/10.1007/S10606-010-9117-8

[23] Sebastian S. Feger, Paweł W. Wozniak, Lars Lischke, and Albrecht Schmidt. 2020. "Yes, I comply!" Motivations and Practices around Research Data Management and Reuse across Scientific Fields. In *Proceedings of the ACM on Human-Computer Interaction,* New York, NY, USA, 1–26. DOI:https://doi.org/10.1145/3415212

[24] Katrina Fenlon. 2019. Modeling Digital Humanities Collections as Research Objects. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL),* 138–147.

[25] Ann Game and Andrew Metcalfe. 1996. *Passionate sociology.* Sage.

[26] Lisa Gitelman (Ed.). 2013. *Raw data is an oxymoron.* MIT press.

[27]   Janet Heaton. 2008. Secondary analysis of qualitative data: An overview. *Hist. Soc. Res.* 33, 3 (2008), 33–45. DOI:https://doi.org/10.12759/HSR.33.2008.3.33-45

[28]   Andrew Iliadis and Federica Russo. 2016. Critical data studies: An introduction: *Big Data Soc.* 3, 2 (October 2016). DOI:https://doi.org/10.1177/2053951716674238

[29]   Helena Karasti, Karen S. Baker, and Geoffrey C. Bowker. 2002. Ecological storytelling and collaborative scientific activities. *ACM Siggr. Bull.* 23, 2 (August 2002), 29–30. DOI:https://doi.org/10.1145/962185.962197

[30]   Karina Kervin, Robert B. Cook, and William K. Michener. 2014. The Backstage Work of Data Sharing. *Proceedings of the 18th International Conference on Supporting Group Work - GROUP '14*, 152–156. DOI:https://doi.org/10.1145/2660398.2660406

[31]   Rob Kitchin. 2014. *The data revolution. Big data, open data, data infrastructures & their consequences.* London: SAGE.

[32]   Rob Kitchin. 2021. *Data Lives: How Data Are Made and Shape Our World.* Policy Press.

[33]   Rob Kitchin and Tracey P Lauriault. 2014. Towards critical data studies: Charting and unpacking data assemblages and their work. In *Thinking big data in geography: new regimes, new research*, Jim Thatcher, Andrew Shears and Josef Eckert (eds.). University of Nebraska Press. Retrieved August 31, 2021 from http://ssrn.com/abstract=2474112http://www.nuim.ie/progcity/

[34]   Cole N. Knaflic. 2015. *Storytelling with data: A data visualization guide for business professionals.* John Wiley & Sons.

[35]   Jintae Lee. 1997. Design rationale systems: Understanding the issues. *IEEE Expert. Syst. their Appl.* 12, 3 (1997), 78–85. DOI:https://doi.org/10.1109/64.592267

[36]   Charlotte Linde. 2001. Narrative and social tacit knowledge. *J. Knowl. Manag.* 5, 2 (2001), 160–171.

[37]   Sara Mannheimer, Amy Pienta, Dessislava Kirilova, Colin Elman, and Amber Wutich. 2018. Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. *Am. Behav. Sci.* 63, 5 (June 2018), 643–664. DOI:https://doi.org/10.1177/0002764218784991

[38]   George E. Marcus. 1994. On ideologies of reflexivity in contemporary efforts to remake the human sciences. *Poet. today* 15, 3 (1994), 383–404.

[39]   Kate McDowell. 2018. Storytelling: Practice and process as non-textual pedagogy. *Educ. Inf.* 34, (2018), 15–19. DOI:https://doi.org/10.3233/EFI-189003

[40]   Niamh Moore. 2006. The Contexts of Context: Broadening Perspectives in the (Re)Use of Qualitative Data: *Methodol. Innov. Online* 1, 2 (January 2006), 21–32. DOI:https://doi.org/10.4256/MIO.2006.0009

[41]   Thomas P. Moran and John M. Carroll. 1996. Overview of Design Rationale. In *Design rationale: Concepts, techniques, and use*, Thomas P. Moran and John M. Carroll (eds.). CRC Press, 1–19. DOI:https://doi.org/10.1201/9781003064053-1/OVERVIEW-DESIGN-RATIONALE-THOMAS-MORAN-JOHN-CARROLL

[42]   Gaia Mosconi, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. Three Gaps in Opening Science. *Comput. Support. Coop. Work* 28, 3–4 (June 2019), 749–789. DOI:https://doi.org/10.1007/S10606-019-09354-Z

[43]   Jessica Mozersky, Heidi Walsh, Meredith Parsons, Tristan McIntosh, Kari Baldwin, and James M. DuBois. 2020. Are we ready to share qualitative research data? Knowledge and preparedness among qualitative researchers, IRB Members, and data repository curators. *IASSIST* 43, 4 (2020).

[44]   Bren Neale. 2013. Adding Time into the Mix: Stakeholder Ethics in Qualitative Longitudinal Research. *Methodol. Innov. Online* 8, 2 (August 2013), 6–20. DOI:https://doi.org/10.4256/MIO.2013.010

[45]   Sharon D Nelson and John W Simek. 2011. Data Dumps: The Bane of E-Discovery. *OR. ST. B. BULL* 71, (2011), 36–36.

[46]   Adegboyega Ojo and Bahareh Heravi. 2017. Patterns in award winning data storytelling: Story types, enabling tools and competences. *Digit. Journal.* 6, 6 (July 2017), 693–718. DOI:https://doi.org/10.1080/21670811.2017.1403291

[47]   Katie Rawson and Trevor Muñoz. 2016. Against cleaning. *Curating Menus* 6, (2016), 1–14.

[48]   Elizabeth Restrepo and Lisa Davis. 2003. Storytelling: Both Art and Therapeutic Practice. *Int. J. Hum. Caring* 7, 1 (February 2003), 43–48. DOI:https://doi.org/10.20467/1091-5710.7.1.43

[49]   Nicole Ruggiano and Tam E Perry. 2019. Conducting secondary analysis of qualitative data: Should we, can we, and how?: *Qual. Soc. Work* 18, 1 (April 2019), 81–97. DOI:https://doi.org/10.1177/1473325017700701

[50]   Anne Ryen. 2011. Ethics and qualitative research. *Qual. Res.* 3, (2011), 416–238.

[51]   Harvey. Sacks. 1992. *Lectures on conversation: Volume I.* Blackwell, Massachusetts.

[52]   Deborah Sole and Daniel Gray Wilson. 2002. Storytelling in Organizations: The power and traps of using stories to share knowledge in organizations. *LILA, Harvard, Grad. Sch. Educ.* (2002), 1–12.

[53]   Brian W Sturm and Sarah Beth Nelson. 2016. With Our Own Words: Librarians' Perceptions of the Values of Storytelling in Libraries. *Storytell. Self, Soc.* 12, 1 (November 2016), 4–23. DOI:https://doi.org/10.13110/storselfsoci.12.1.0004

[54]   Andrea K Thomer and Karen M Wickett. 2020. Relational data paradigms: What do we learn by taking the materiality of databases seriously? *Big Data Soc.* 7, 1 (June 2020). DOI:https://doi.org/10.1177/2053951720934838

[55]   Andrew Treloar. 2008. Data management and the curation continuum: how the Monash experience is informing repository relationships. *VALA2008 Proc.* (2008), 13. Retrieved from http://www.vala.org.au/vala2008-proceedings/vala2008-session-6-treloar

[56]   Alexander C. Tsai, 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. *Soc. Sci. Med.* 169, (November 2016), 191–198. DOI:https://doi.org/10.1016/J.SOCSCIMED.2016.08.004

[57]   Nadia De Vecchi, Amanda Kenny, Virginia Dickson-Swift, and Susan Kidd. 2016. How digital storytelling is used in mental health: A scoping review. *Int. J. Ment. Health Nurs.* 25, 3 (June 2016), 183–193. DOI:https://doi.org/10.1111/INM.12206

[58]   Janet Vertesi and Paul Dourish. 2011. The Value of Data : Considering the Context of Production in Data Economies. *Cscw2011* (2011), 533–542. DOI:https://doi.org/10.1145/1958824.1958906

[59]   Janet Vertesi, Jofish Kaye, Samantha N. Jarosewski, Vera D. Khovanskaya, and Jenna Song. 2016. Data Narratives: uncovering tensions in personal data management. *Proc. 19th ACM Conf. Comput. Coop. Work Soc. Comput. - CSCW '16* (2016), 477–489. DOI:https://doi.org/10.1145/2818048.2820017

[60]   Jillian C. Wallis, Elizabeth Rolando, and Christine L. Borgman. 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS One* 8, 7 (July 2013), e67332. DOI:https://doi.org/10.1371/JOURNAL.PONE.0067332

[61]   Peter Walters. 2009. Qualitative archiving: Engaging with epistemological misgivings. *Aust. J. Soc. Issues* 44, 3 (March 2009), 309–320. DOI:https://doi.org/10.1002/J.1839-4655.2009.TB00148.X

[62]   Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 1 (March 2016), 1–9. DOI:https://doi.org/10.1038/sdata.2016.18

[63]   Ann Zimmerman. 2007. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *J. Digit. Libr. 7, no. 1-2 (20* 7, 1–2 (2007), 5-16.

[64]   Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni, Ronald Meijer, and Roexsana Sheikh Alibaks. 2012. Socio-technical Impediments of Open Data. *Electron. J. e-Government* 10, 2 (December 2012), 156–172. Retrieved September 1, 2021 from https://academic-publishing.org/index.php/ejeg/article/view/571

[65]   Anneke Zuiderwijk and Helen Spiers. 2019. Sharing and re-using open data: A case study of motivations in astrophysics, International Journal of Information Management. *Int. J. Inf. Manage.* 49, (2019).