Fostering Research Data Management in Collaborative Research Contexts: Lessons learnt from an 'Embedded' Evaluation on designing a 'Data Story'

Gaia Mosconi (gaia.mosconi@uni-siegen.de)
University of Siegen
Aparecido Fabiano Pinatti de Carvalho
University of Oslo
Hussain Abid Syed
University of Siegen
Dave Randall
University of Siegen
Helena Karasti
Volkmar Pipek
University of Siegen

Research Article

Keywords: Research Data Management, Research Collaboration, Design, Embedded evaluation, Practice-centred research

Posted Date: November 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2255943/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Computer Supported Cooperative Work (CSCW) on May 15th, 2023. See the published version at https://doi.org/10.1007/s10606-023-09467-6.

Abstract

Research Data Management (RDM) practices – spanning the curation, storage, sharing, and reusability of research data – are foundational to the Open Science (OS) agenda. Because of that, many and various funding institutions are increasingly demanding active engagement of researchers in them. Nevertheless, recent studies suggest that RDM practices are not yet properly integrated into daily research workflows, nor supported by any tools researchers typically use. In this paper, we introduce evaluation results of a technological aid for implementing a design concept called 'Data Story', drawing on ideas from digital and data storytelling. This concept has been devised to facilitate the appropriation of RDM practices by researchers working mainly with qualitative data in their daily workflows. It integrates traditional data curation approaches with a more narrative, contextual, and collaborative organizational layer that can be thought of as a 'story'. Our findings come from a long-term 'embedded' evaluation of the concept and show: (1) the potential benefits of engaging with a Data Story for RDM; (2) Data Curation issues and learning opportunities; and (3) a broader set of issues and concerns that remain unaddressed in the current state of play. Our contribution, based on lessons learnt, is to provide a new design approach for RDM and for new collaborative research data practices, one grounded in narrative structures, capable to negotiate between top-down policies and bottom-up practices, which supports 'reflective' learning opportunities - with and about data - of many kinds.

1 Introduction

Problems related to collaborative practices are frequently related to 'infrastructural' work that may well benefit some practitioners, or a community as a whole, but not the practitioners who need to do the work (Grudin 1988). In those cases, these 'beneficial' rules and procedures are often well-known and acknowledged, but their appropriation into actual practices often proves difficult. This also applies to research contexts, where, in principle, the Open Science (OS) agenda can provide a beneficial framework for successful collaborations. In fact, the Open Science mandate, strongly supported by funding and research agencies who aim to facilitate research verifiability, 'good' scientific practices, and data reuse, is simultaneously changing the dynamics of research (Wallis et al., 2013) and promoting massive infrastructural investments. This movement has been conceptualised, within the FAIR (Findability, Accessibility, Interoperability, and Re-use) data principles, as entailing guidelines to improve Research Data Management (RDM). This has been realised in an ever-increasing proliferation of data hubs and repositories acting as storage and recovery media in research (Borgman et al., 2019; Wilkinson et al., 2016). The top-down policy-driven adoption of OS initiatives is often constrained due to funding agencies' insistence on a generic view of research data practices, and a strong emphasis on data storage and recovery as the primary issue. However, more recently, concerns for how data is to be understood across disciplinary boundaries, and how re-use is to be facilitated, have come to the fore (Feger et al. 2020). In turn, this implies that discipline and methodological-specific norms and data practices need to be investigated and understood (Borgman 2012, 2015; Pasquetto et al., 2016; Velden 2013).

RDM, in itself, is a complex and long-term endeavour spanning the entire research lifecycle and beyond, requiring attention to the specifics of data creation, curation, storage, sharing and reusability (Treloar and Harboe-Ree, 2008; Whyte and Tedds, 2011). They are different practices but at the same time intertwined. 'Good' RDM asserts the notion of reusability through openness, sharing and collaboration throughout the whole research process but the implications for RDM when confronted with disparate data practices applied by different disciplines, methodologies, and research communities are still not fully understood. For example, in Humanities and Social Sciences (HSS), collaborative and data-intensive research endeavours, the plurality of research methods, standards and traditions, ethical and legal implications, and heterogeneous practices in storing, processing, sharing and analysing data indicate higher barriers to the implementation of OS initiatives (Eberhard and Kraus, 2018; Korn et al., 2017; Mosconi et al., 2019). Another layer of complexity in RDM is added by the overhead (additional work, time, and costs) implied in the appropriation of data curation and the sharing practices which require researchers to engage in systematic organization of data (i.e.: metadata creation, contextualization and structuring the storage of data) in on-going research projects and in anticipation of reuse, verifiability, and collaboration.

To tackle some of these complex problems, new tools, and research data infrastructures for RDM are emerging (Borgman et al., 2019; Kaltenbrunner 2017; Khan et al., 2021; Lee et al., 2009). In our view, these solutions typically address the guidelines of findability and accessibility, but they do not necessarily solve the issue upstream of how to curate and manage data effectively during the research process. It is clear that tools for the meaningful appropriation of RDM as a long-term processual phenomenon are as yet lacking. Here, we argue, data storytelling approaches can come in handy. Over the past few years, data storytelling – i.e., the use of narrative and visual elements to effectively communicate data insights (Dykes 2015) – has been emerging as "a promising approach for supporting more accessible and appealing human-data-interactions" (Concannon et al., 2020, p. 2). However, as we will argue in section 2, little has been done to support researchers working in an interdisciplinary context to use such an approach to manage, share and potentially re-use data. Our current contribution specifically addresses this gap and seeks to provide conceptual and socio-technical answers to some of the issues above. Our work is, therefore, driven by the following question: How can we best support the appropriation and the establishment of RDM practices of researchers – mainly working with qualitative and ethnographic data – in collaborative research contexts?

Since 2016, we have explored socio-technical contexts in which qualitative-ethnographic data are produced, curated, and eventually shared. These insights allowed us to delineate the gaps that still exists between the OS and RDM top-down agenda and the bottom-up practices of researchers affected by it (Mosconi et al., 2019). These investigations have been carried out within an information management (INF) project, connected to a Collaborative Research Centre (CRC), and funded by the German Research Foundation (German acronym: DFG), where the DFG expects INF to provide support and develop RDM solutions for the qualitative and ethnographic-oriented research projects (representing the majority in our CRC).

Driven by these institutional constraints and drawing on empirical findings, we developed a conceptual solution for RDM called ‘Data Story’ (Mosconi et al., 2022) which offers a means of enhancing and naturalizing curation practices through storytelling. The name itself *Data Story* is not new. We credit the term to Nancy Duarte (2019) who has been applying data storytelling principles to support decision-making processes within the business sector. The novelty here, however, lies in the application of data storytelling insights to the field of RDM and in the use of the ‘Story’ as a metaphor and design principle used to implement a socio-technical system in support of RDM practices not yet established.

Adding to our previous conceptually oriented publication (Mosconi et al., 2022), this paper reports on the design as it was iterated, based on user interaction and feedback. We define our engagement and evaluation as ‘embedded’ – (see e.g., Barry et al., 2018; Lewis and Russell, 2011) meaning that researchers and research participants are ongoingly immersed in the research context in which the technology is to be used. In this way, ‘Data Story’ became both the topic and the medium through which we were able to understand how RDM practices can be introduced into researchers’ daily workflows and how collaborative research contexts can profit from them. Our contribution, based on lessons learnt, is to provide a new design approach for RDM and for new collaborative research data practices, one grounded in narrative structures, capable of negotiating between top-down policies and bottom-up practices, and which supports ‘reflective’ learning opportunities - with and about data - of many kinds.

2 Related Work

Adding some form of narrative to data forms and structures has been advocated and implemented in a variety of contexts. This can be seen, for instance, in the literature on ‘digital storytelling’ and on ‘data storytelling’.

Previous research has investigated the use of storytelling in non-profit organisations (Erete et al., 2016) and in educational contexts – e.g., (Martinez-Maldonado et al., 2020; Xu et al., 2022). The InfoVis community, as it is sometimes termed, has invested considerable effort in providing tools for generating effective visualisations to aid narrative - see e.g., (Fekete 2004; Fekete et al., 2008; Liu et al., 2014; Méndez et al., 2017; Pantazos and Lauesen, 2012). Recently, some attention has been paid to the differences in meaning that users in different contexts might experience (Lallé and Conati, 2019). This latter issue is of central importance to our own work. Elsewhere, ‘digital storytelling’, as it is sometimes called, has explored the use of visuals, for instance, in education (Wu and Chen, 2020), health (Moreau et al. 2018; West et al. 2022), and in the business sector (Duarte 2019; Knaflic 2015). However, very little work has been done to support researchers working in an interdisciplinary context to use data storytelling insights to manage, curate, share and potentially re-use data. Even less work has used such insights to develop socio-technical solutions for RDM issues and related practices.

2.3 Challenges for RDM: the issues with Data Curation and Sharing

Research Data Management (RDM) is commonly defined as "the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results" (Whyte and Tedds, 2011, p.1). RDM is characterised by several core practices, such as data curation, metadata documentation, long-term preservation, and data sharing altogether leading to the publishing and successful reuse of research data.

Ethical issues, privacy concerns, technical limitations, lack of skills, restricted access, and lack of a rewards systems are among the most discussed barriers to effective RDM in all major disciplines and fields (Feger et al., 2020). In fact, curating, preserving, and sharing research data require appreciable overhead and technical skills but the current scientific culture and rewards system do not directly incentivise or yet, recognise these endeavours (Fecher et al., 2017). Moreover, issues in sharing data are intrinsic to the complex and contextual nature of data itself. Data are not 'natural kinds' but are constructs, existing in contexts of production, use and reuse (Borgman 2015).

Nonetheless, some disciplines, such as the natural sciences, have managed to adjust better to OS and RDM expectations, and progressively, have developed internal policies to ensure the curation, sharing and eventually reuse of research data (Zuiderwijk and Spiers, 2019). For other disciplines these requirements are relatively new, and researchers and institutions are still struggling to understand how to meet these new demands.

For Humanities and Social Sciences (HSS), and specifically for those researchers working with qualitative data, the expectations for data curation and sharing pose some additional challenges characterised as epistemological, methodological, and ethical in nature (Feldman and Shaw, 2019; Ryen 2011). For instance, with these data, legal and ethical issues can abound, 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 data and/or metadata, and the sheer heterogeneity of RDM practices can make standardisation massively problematic. Therefore, data sharing concepts and infrastructures for quantitative data cannot be translated directly to qualitative data. As Tsai et al. (2016) puts it

"… the iterative nature of qualitative data analysis, and the unique importance of interpretation as part of the core contribution of qualitative work, [makes data] verification likely to be impossible" (p. 192).

Other critical factors are the protection of study participants that might be imposed by research ethic bodies, or self-imposed through researchers' lack of familiarity with ethical data sharing practices involving human subjects, but also trust-related issues: researchers lack the knowledge on who might have access to their data once shared and what they will do with it, fearing in the long-term to lose control over the data and maybe even endanger study participants (Eberhard and Kraus 2018). Another pressing problem is that, for the most part, only major universities, libraries, and librarians are the service providers for RDM support and training. These institutions are often understaffed or unqualified to advise on a huge variety of disciplines and heterogenous research data practices (Hamad et al., 2021; Kervin et al., 2014; Pinfield et al., 2014). Therefore, they might fail in satisfying the increasing demand for skills in RDM applied in different research contexts.

It is evident that data curation and sharing still has unresolved and nuanced challenges. In our contribution, we seek to address some of the abovementioned issues by examining a solution that is innovative, flexible, epistemologically nuanced, and which has been designed by closely looking into situated, collaborative research data practices.

2.2 Existing Solutions and Infrastructures for RDM, Data Curation and Sharing

Some major barriers to the appropriation of data curation and sharing practices can be rooted in the interaction with socio-technical infrastructures or in the lack of suitable ones (Borgman 2010; Edwards et al., 2013; Feger et al., 2020). Most existing solutions are repository-styled research storage facilities: they can be generic, such as Zenodo[1], Dryad[2] or DataverseNO[3], supporting many types of research data and therefore suitable for a wide variety of scientific fields; or they can be discipline specific and community-driven, e.g., for social science research, examples being QualiService[4], GESIS[5], and SowiDataNet[6] (Linne and Zenk-Möltgen, 2017). Universities' repositories are also being increasingly developed by all major institutions, and they often address multiple disciplines similar to existing generic repositories.

Research repositories, however, target two specific aspects of the RDM data life cycle: long-term preservation and sharing. They do not necessarily solve the issue upstream on how to curate and manage date effectively during the research process (Mosconi et al., 2019). Archiving data in a repository is then seen by researchers as the ultimate step, not directly connected to daily practices in which data get generated, processed, and analysed, causing the archiving process to be perceived simply as an extra burden, with no direct benefits, especially in the absence of a strong mechanism of rewards (Chawinga and Zinn, 2020; Curdt and Hoffmeister, 2015; Donner 2022).

Moreover, 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 these 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 data objects (Feger et al. 2020; Mosconi et al., 2019). These previous studies shown that lack of suitable infrastructure, knowledge and skills has forced researchers to adopt haphazard, ad hoc, practices that lead to unstructured archives.

A thorough understanding of RDM in practice is clearly indicated if, as Feger et al. (2020) suggest, HCI research is to have a role "in supporting the transition to effective digital RDM through a design-focused understanding of the roles and uses of technology". Our prior work on the use of data stories in the context of RDM (Mosconi et al., 2022) has demonstrated at a conceptual level the potential role of narrative structures in providing relevance for data curation and sharing. However, only a very limited amount of work has been aimed at innovative digital solutions which address these problems (Feger et al., 2019; Garza et al., 2015; Mackay et al., 2007). One notable example for the Humanities is PECE (worldpece.org), an open-source, Drupal-based platform designed to support a wide range of collaborative humanities projects, which pays a considerable attention to the way data artefacts get collaboratively shared, archived, and potentially reused (Fortun et al., 2021; Poirier 2017).

2.3 Existing recommendations for the design of RDM tools and infrastructures

Recent literature has identified design recommendations for new tools and infrastructure in support of RDM (Feger et al., 2020; Koesten et al., 2019), and more specifically for data curation and sharing (Birnholtz and Bietz, 2003; Feger et al., 2019; Jahnke and Asher, 2012; Rowhani-Farid et al., 2017; Zimmerman 2007). Because these two practices (data curation and sharing) directly imply the additional work needed to make data understandable for a potential audience, they are often described in relation to reuse.

For instance, Koesten and Simperl (Koesten and Simperl, 2021) argue that in order to better facilitate reuse, the creation of structured textual data documentation (or descriptions such as Readme files) are of importance, as they often constitute the first points of interaction between a user and a dataset. Therefore, their creation should be supported during the act of curation and sharing. As they put it:

we cannot see datasets as usable end products without telling the story of how they were made. Because the story is complex, the user experience of data relies on tools and environments that try to do exactly that: embedding datasets in the rich context of their creation and use (Koesten and Simperl, 2021, p.99)

Other studies (Birnholtz and Bietz, 2003) underline how research infrastructures also need to improve communication channels around research artefacts because anything that is shared should in principle be of interest for somebody else and data creator and recipient need to be allowed to exchange information. Rowhani-Farid et al. (2017) and Feger et al. (2021), on the other hand, concentrated on tools for sharing and reproducibility and stressed the importance of mechanism of reward, to increase motivation and benefit, which could be promoted through OS badges and gamification elements.

Technical standards, legal frameworks, and guidelines are also crucial and need to be considered while designing new tools and infrastructure but most of the literature in this direction has focused on operational problems such as interoperability and machine readability and not so much on readable metadata for human interpretation. Only a few solutions have been proposed so far to document data context beyond what is typically considered and stored as metadata (Gebru et al., 2021; Preuss et al., 2018).

Feger et al. (2020) suggest investigating how RDM tools could compensate for the lack of formal training in RDM and state that new tools should be developed to remove current barriers and more specifically to integrate RDM practices into the research workflow. In our view, RDM, metadata, and curation work have focused too much on interoperability and machine readability. The issue here for us is how do we produce a meaningful (possibly asynchronous and distant) interaction between users in and through the data they use. In what follows, then, we describe the iterative process by which we designed and evaluated a new technological aid, called 'Data Story', devised to provide for meaningful organisation, curation and sharing of heterogenous data which in the long-term could include all the above suggestions and recommendations made by previous studies.

[1] https://zenodo.org/
[2] https://datadryad.org/stash
[3] https://dataverse.no
[4] https://www.qualiservice.org/de/
[5] https://www.gesis.org/en/research/research-data-management
[6] https://www.re3data.org/repository/r3d100011062

3 Methodology And Approach

In this section, we describe the ethnographic, long-term (and ongoing) engagement taking place within the aforementioned information management project (INF). This involvement has *inter alia* produced the Data Story design concept. This concept, as introduced above, was meant to support researchers to engage in data storytelling as a way to curate qualitative ethnographic data to be shared with other researchers.

In order to enhance the likelihood of designing a useful and usable concept, which can be integrated in current research data practices and appropriated accordingly, we drew on a practice-centred approach predicated on constant engagement with the user and their contexts (Wulf et al., 2015). Therefore, the interests and concerns of all parties guided our interaction in the field, and continuously shaped our design and evaluation activities from within.

Evaluation, of course, can take many forms. It can be conceived of, for brief mention, as 'summative', 'formative', 'diagnostic', 'situated', and so on (Chambers 1994; Irani 2010; Kaye 2007; Ledo et al., 2018; MacDonald and Atwood, 2013; Remy et al., 2018; Twidale et al., 1994). The character of each is shaped by epistemological assumptions, pragmatic considerations, and overall purpose. As briefly mentioned above, our evaluation can be described as 'embedded' (Lewis and Russell, 2011) due to the nature of our participation, which is long term, involves ongoing interaction with participants, is participative but at the same time constrained, meaning that the aims of all participants are restricted by the institutional framework and expectations. An important element of this is that there are no obvious demarcations between investigative, design, and evaluative work. All can be seen as being mutually constitutive. Below, we provide details concerning the research context and the design work as it unfolded through our engagement and report on the major evaluation activities which shaped our design: formal and informal meetings, thinking aloud evaluation sessions, a focus group and follow-up interviews.

3.1 Research context: Long-term engagement in a Collaborative Research Centre

The CRC[7] is composed of 14 projects with over sixty researchers, representing several major disciplines and faculties, and where the majority of them apply qualitative and ethnographic methods. As expected by our funding agency (DFG) and defined by the project proposal, the goal of the INF project is to develop (and establish) RDM practices and infrastructural solutions which should lead to the curation, sharing, and potential reuse of research data. Since September 2016, the first author has been investigating the difficulties of qualitative data sharing and the practical challenges that the OS agenda is presenting specifically in qualitative-ethnographic driven research contexts (Mosconi et al., 2019). She has been collaborating with the IT service provider of the University, helping developers to customise several open-source tools (i.e.: *RDMO*: for creating Research Data Management plans; *DSpace:* a long-term repository; and *Humhub,* a platform for team collaboration and sharing). In particular Humhub, which is now named 'Research-hub', was established to customise, test, and study new RDM concepts and workflows. These are expected to be implemented by INF in the long-term. In parallel, she has conducted over thirty qualitative interviews and ethnographic observations, run meetings to discuss RDM issues with CRC's projects, and supported them in creating their RDM plans.

With our first interviews and observations, conducted between 2017 and 2019, we investigated researchers' data life cycle (with a particular focus on sharing and curation practices), and their issues with socio-technical infrastructure. Our initial insights allowed us to discover major gaps that still exist between the OS *grand vision* and the bottom-up research data practices observed in the field (Mosconi et al., 2019). It was evident that, while sharing and curation practices are expected by all major funding agencies, these practices are not yet supported by any tool that researchers use daily, nor they are integrated in researchers' workflows. If at all, they are performed informally or in a haphazard way. Consequently, as already highlighted by previous literature (Begley and Ellis, 2012; Collaboration 2012; Fecher et al., 2017), data curation and sharing practices, needed to meet the Open Science goals, are perceived by many as an unrewarding chore. Put another way, their primary work tasks are typically separate from any additional work they might need to perform for others to benefit. In the context of data curation and sharing, the beneficiaries are, or appeared to be, mainly future (unknown) data re-users. Indeed, much of the scepticism about the funding agency's agenda that we encountered early on in our work was a function of these factors. Others, however, showed an interest in innovative solutions that might help them to represent and share their highly heterogenous research data, initially for their own purposes. They were specifically interested in how to organise different data sources and underpin the work of collaborative interpretation and sense-making. These early investigations led us to envision a system called Data Story (Mosconi et al., 2022), in which researchers could organise portions of pre-selected data to be curated with written narratives, storytelling, tags and metadata elements, ultimately to share them with colleagues and/or with an external audience. We anticipated that, in the long-term, the Data Story would help to introduce and support the new RDM practices expected by the DFG.

3.2 Data Story design rationale: Sketches and low-fidelity prototype

The concept was inspired by the way researchers were seen to share ‘data snippets’ and engage with them on an ad hoc basis during meetings, collaborative analyses sessions or paper discussions (for more details, see Mosconi et al., 2019, 2022) In those meetings, portions of selected data are contextualised to others with the support of written or oral narratives and collaboratively interpreted and analysed. Through collaborative research data practices, 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 mobilised as part of broader processes of interpretation and meaning-making” (p.1). Therefore, the main rationale behind the concept was to allow the sharing of heterogenous qualitative data accompanied with 1) written narratives or storytelling practices for data contextualization, analysis, and sense-making; and 2) technical element and standards, such as metadata, tags and DOI for data curation and retrieval.

Initial prototype sketches were made between January and February 2021. Figure 1 shows the Data Story as an independent module already integrated and accessed through the Research-hub platform menu (already established in 2019).

We took the story as a design metaphor and organizing principle and as such, we translated this into ‘design features’ that would reflect a Story-like structure. Therefore, we organised its interface with chapters and a panel that would allow movement across them. The sketches developed further into a low-fidelity prototype designed between February and March 2021.

To simplify the possibilities, we created three main chapters: 1) project set-up; 2) data processing; 3) findings (see Figure 2 below: Data Story overview). Open text fields for writing narratives, tags, relevant metadata and a DOI were organised all along the three interface chapters. Especially in the data processing chapter, researchers would showcase pre-selected data, organised them in sub-sections, and visualised them along a timeline. To better support the data creators in engaging with narrative and storytelling practices, we highlighted relevant guiding questions called ‘tips’ next to each open text field, that researchers would use to structure their stories and contextualise their data. Finally, we envisioned a plugin for different tools (i.e., Word, Sciebo, Maxqda etc.) that would allow researchers to easily add new data to their stories ‘on the go’ while still actively working on their research projects.

In the next section, we provide details concerning the evaluative work we conducted and illustrate how the prototype changed accordingly and how progress was made on the wider question of supporting RDM and collaborative research data practices.

3.3 How the ‘Embedded’ Evaluation shaped the design

As mentioned above, our overall ethnographic approach is characterised by a long-term engagement and by member participation, while the type of evaluation conducted can be described as ‘embedded’ (Lewis and Russell, 2011), meaning that evaluation opportunities spontaneously emerged from our double role and our ongoing engagement in the field. In fact, since 2016 we have been members of the CRC ourselves, so we are part of the context we were called to design for (and with). We always positioned ourselves in a constant dialog with the researchers involved whom we met regularly during informal encounters, official plenary meetings, and seminars organised by ourselves or others in the CRC.

As showed in Figure 3, initial brainstorming and the low-fidelity prototype were grounded on previous interviews and observations, while evaluation of feedback on our conceptual design was done initially in a PhD forum (May 2021, with twelve participants), and in a strategic planning meeting locally known as 'Retreat' (July 2021) where all CRC's projects (including our own) were invited to discuss their latest updates concerning publications and research findings. On both occasions, the first author shared with the participants the low-fidelity prototype and the draft of a conceptual paper which described it. Researchers were enthusiastic with our initial concept, with our interpretation of their RDM issues, and with the new opportunities that a Data Story could offer. As one PhD student told us:

I really like the idea of combining few metadata and organised the data and information across the research process that you divided in chapters. I like the fact that you could use a Data Story over time and add more data to it. In this way, you could use the interface to discuss relevant data with your colleagues and even with others who do not directly work with you. (PhD forum, May 2021; PhD Student in HCI)

Another Postdoc said during the Retreat:

Data Story could be used to collaborative craft publications outcomes based on specific relevant data but also as a possibility to present to a wider audience how data practices actually unfold. I find this approach very exciting. I really want to use it at some point to see how it works. (Retreat, July 2021; Postdoc in Media History)

3.3.1 Thinking Aloud evaluation sessions

After this initial positive feedback, we decided to evaluate the prototype workflow in the actual interface of the Research-hub platform where the Data Story is planned to be fully implemented. We especially wanted to find out what researchers liked or disliked about our design, how they would engage with its workflow, what was missing or unclear, and what further ideas or expectations researchers might have. We then designed a high-fidelity prototype that mimicked the Research-hub platform interface but with the same features and structure of the low-fidelity described in section 3.2. With it, we ran six individual thinking aloud evaluation sessions between July and August 2021. Three graduate students and three Postdocs representing all major disciplines were invited to join the sessions via Zoom (see Table 1 for participants' overview). Each participant received the clickable prototype[8] link at the beginning, then the first author instructed them to share their screens, engage with the Data Story workflow and provide feedback by thinking aloud (Van Den Haak et al., 2003).

Table 1: Participants overview: background, role and type of evaluation performed with them. All participants have an interdisciplinary background and apply qualitative and ethnographic methods in their research with various degree of expertise.

ID	Pseudonym	Main Background	Academic Role	Type of Evaluation	Date
#1	Claudia	HCI	Ph.D.	Think Aloud Ev. Session	13.09.2021
#2	Oliver	Media history	Postdoc	Think Aloud Ev. Session	05.08.2021
#3	Karl	Computer Science	Postdoc	Think Aloud Ev. Session	06.08.2021
#4	Paul	STS and Media Studies	Ph.D.	Think Aloud Ev. Session	06.08.2021
#5	Rose	Economics	Ph.D.	Think Aloud Ev. Session	16.08.2021
#6	Marie	Educational Science	Postdoc	Think Aloud Ev. Session	24.08.2021
#7	Alex	Software Engineering	Ph.D.	Focus group + Interview	20.01 + 10.02.2022
#8	Franziska	Media Science	Ph.D.	Focus group + Interview	20.01 + 25.02.2022
#9	Dave	Computer Science	Master	Focus group	20.01.2022
#10	Max	Sociology	Postdoc	Focus group + Interview	20.01+15.02.2022

The initial feedback, collected in the PhD forum and Retreat, were enthusiastic and positive. However, when confronted with the first prototype, researchers were more critical, and some scepticism was again expressed. Researchers were especially discouraged by the amount of metadata and input fields distributed across all sections. They spotted some redundancies concerning metadata and tags, and they found some metadata confusing and difficult to fill in. In general, they were confused with the purpose of a Data Story in the first place and wondered why one would put to so much effort into it.

Based on this feedback, we modified the prototype and created a second version[9] with fewer sections and less metadata. We removed the option to provide metadata for single files and focused the design on open narratives and open input text fields. As shown in the Figure 4, the prototype lost the rigid chapter structure but maintained the timeline of data and related methods. More emphasis is given to the narrative itself, data, and methods, to be described with open text fields.

3.3.2 Focus Group and follow-up interviews

During all evaluative activities, participants mentioned repeatedly how they missed the opportunity to engage with the actual writing flow, they were concerned with how long that would take, and how a Data Story would look like in the end. Therefore, we organised a focus group to discuss specifically the writing process and with the goal of creating the first sample of users Data Stories. The focus group was organised around two solo-writing timeslots (40 min each) and two plenary discussions timeslots (45 min each). Four different participants were invited this time (see Table 1 for overview). Researchers were invited to selected beforehand a few sample data (pictures, interviews, surveys etc.) that they collected during their research project and that they imagined sharing with an external audience via the Research-hub platform.

At the beginning of the workshop, we briefly introduced the Data Story concept and showed the high-fidelity prototype. We created an online form with the tool Tripetto to collect and save all written stories and sample data uploaded by the participants. After the focus group, we copied and pasted all stories and data researchers uploaded (via Tripetto) into the new interface design. We also included social media features, such as likes and comments to provide a real feeling of the potential interactions. Finally, we had one-hour follow-up interview with the focus group participants to discuss the Data Story visualization and interface navigation. One week after the follow-up interview, one of the participants came back to us a with the following feedback:

This has been fun. I made some reference to the tool at today’s meeting on the annual conference because we were talking about the need for new forms of presentation (actually, also briefly discussed the upcoming Retreat). I guess there’s plenty of interest, at least on the doc/postdoc level (email sent by Max, a Postdoc, to the first author).

3.4 Data collection and analysis

All interactions mentioned up to this point – the PhD forum, Retreat, thinking aloud evaluation sessions, focus group (plenary sessions), follow-up interviews – took place via zoom due to the pandemic restrictions. They were all video recorded and transcribed ‘ad litteram’. For all the other informal interactions, meetings, or seminars we wrote fieldwork reports. The thinking aloud evaluation sessions and the follow-up interviews lasted in average 1 hour.

After repeated reading, all data were open coded (Strauss and Corbin, 1998), and structured into approximate categories organised into similar statements that reflected the issues raised by the respondents. Iterative data analysis sessions took place between September and October 2021 (for the thinking aloud evaluation sessions) and between February 2022 and April 2022 (for all data combined). The first author, as data collector, was leading the sessions. In the very first analysis sessions, the first author and more experienced researchers met to discuss, adapt, and sometimes align the emerging themes, following a broadly inductive analytic procedure (cf. Thomas 2006). The broader categories that emerged from the analysis were: 1) personal and collaboration benefits connected to sharing data, 2) RDM issues and expectations, 3) open issues and fears. The first author expanded those themes and checked for inconsistencies.

The focus of the evaluation and analysis was not on the tool or the interface itself, but rather on what we had learned through this evaluation process concerning how to foster new research (management) practices. The focus was on how to analyse the way in which researchers reasoned about how to think, select, describe, and write about data when engaging with the Data Story, and what issues emerged in doing so. The ongoing evaluation, then, was critical to our emerging understanding of how to foster RDM practices in collaborative research contexts. It enabled us, simply, deeper into researchers' expectations, hopes, and fears.

[7] CRCs can be funded for up to twelve years across three separate evaluation stages (Phase I; Phase II and Phase III). Our CRC started in January 2016 and completed its first funding period in December 2019. A second phase began in January 2020 (funded until December 2023). All CRC's projects are interdisciplinary in nature.

[8] The version of the low-fidelity prototype can be accessed here: (https://bit.ly/3ry9mH2).

[9] The second version of the high-fidelity prototype can be accesses here: (https://bit.ly/3ehmFEN)

4 Findings

In this section we report on the findings concerning the above-mentioned research question. The first section highlights the benefits that researchers hoped to get from a tool like the Data Story, and stresses those benefits connected to sharing and collaboration research (data) practices. The second section explores issues concerning metadata and curation work while pointing to how researchers could increase their awareness and learn to do this type of work through Data Story. The last section digs deeper into general issues or open questions and explores some anticipated issues that researchers talked about when imagining a Data Story becoming commonplace in academia. Each of those sections is an important building block of the overall answer to our research question. The implications of this are discussed in section 5.

4.1 Identified benefits for research collaboration and sharing

In the focus group, participants engaged in an animated discussion and spelled out several benefits and concrete use cases in which a Data Story could be helpful. For example, Max mentioned how he sees a lot of value in the concept, in the data contextualization and visualization suggested in the prototype. He hoped, for instance, that it might replace the sharing of long papers in CRC's meetings, such as the Retreat and research forums, because in the end "nobody reads papers in detail for lack of time". Data Stories, thus, provide a quick entry point into ongoing collaborative research projects where authors can explain essential information and even display relevant data like interviews or observations. As Max put it:

… it can open opportunities for different discussions and different type of questions to be asked in plenary meetings. […] it forces you to write the essential and test if others understand what you want to say and what your aims are (Focus group, January 2022; Postdoc in Sociology).

In general, participants saw benefit in the time they spent in "sitting with their data" which was useful to them for structuring the major insights of their research process while also having a format specifically targeted to show these insights to others.

Peer learning opportunities were also highlighted. In fact, Alex graduated in software engineering and when he joined an HCI department few years ago, he struggled in adapting to the new research environment. He joined an already existing project that had started two years earlier. Data collected from other colleagues were not accessible, so it was even harder to understand what had been done until that point, to learn from others and/or to start analysing materials already collected from other graduate students. If Data Stories had been available when he joined, he said, he might have had the chance to learn faster how the HCI and CSCW communities deal with data, which methodologies are applied and how. Franziska had a similar experience. She started her project one year later and she needed the overview of what they did before her time, so she decided to visualise her own data in order to get an overview and prompt discussion with other colleagues:

We created a lot of data, and it was also difficult for myself to have an overview. I also visualised it. I discussed the visualization with my colleagues from the other faculty because they didn't know everything that was happening, so it was very good to discuss it together and we used it also as a basis for writing papers just to know what kind of data do we have, what kind of insights did we get (Follow-up interview, February 2022, Ph.D. student in Media science):

In general, participants highlighted the need for an overview and data organization which many of researchers struggles to have and are in constant search of tools or new methods to visualise what has been done collaboratively. Franziska added that she has been searching for quite some time for a tool where to present their results to the funding agency, as a way to provide them with a quick overview of their data collection and research achievements. The Data Story is fitting this specific need, where links to stored data folders could be established, to prove that data exists somewhere, and they are stored safely. Others envisioned Data Stories to be used as prop to collect data in the field, inviting participants for example to create their Data Stories and collaboratively gather data. This is a need that was expressed by one CRC's project where researchers interested in 'decolonizing ethnography' (Bejarano et al. 2019) have been searching for tools where participants could be engaged from the beginning in the data collection to support researchers' claims.

Finally, others stressed the impact that Data Stories could have in the long-term, specifically for re-use or for guiding new line of research and research questions. As Oliver puts it:

Funders want research data to be collected and archived and the question is 'where would it be?' Should I put them on an anonymous archival environment and then it's there for eternity? Or wouldn't we have to invent new formats of decentralised devices connected through the DOI, so that the published texts are somehow connected to their materials?' (Thinking aloud session, August 2021, Post Doc in Media History).

In fact, 'anonymous', remote archives, which are removed from where data are actually created, are often perceived as an additional burden and researchers do not see a benefit in archiving data there. Data Stories instead emphasise the organization, the overview, and analytical insights that researchers want to get from 'their' data, initially for themselves, and later, potentially, provide it to others.

4.2 Data Curation and Metadata Issues

The first high-fidelity prototype integrated technical elements such as the tags, metadata and DOIs to support data curation and retrieval along with open text fields for open contextualization and narratives. However, during the thinking aloud evaluation sessions, researchers found it surprising but also confusing to see these technical elements. For example, Rose was confused, because she wasn't sure what metadata really are and what purpose they might have in the process. As mentioned in section 3, after these feedbacks from the focus group the prototype was redesigned. The majority of metadata were removed, we left the categories more open-ended, and we almost lost the 'traditional' curation aspect. However, in one of the plenary sessions we discussed the issue of standardization which can be connected to the role of metadata. Researchers agreed that standardization would make the process faster and could help in mapping the major methodologies used within a specific research group but also it could generate internal discussions concerning the development of methods by showing in-depth descriptions and sample data that could be compared and might trigger new research collaboration. Max suggested having a workshop in the CRC where together researchers could come up collectively with their own metadata and categories starting with their methodologies and research interests. A couple of researchers also mentioned some metadata elements that could be added as a way of organizing and detailing the data. For example, for interview data they mentioned "place of interview, date of interview, length of interview etc." as metadata that could be helpful to describe single data items and organizing them along the timeline. Interesting to note is that on a different occasion, after a seminar organised on the topic of RDM and curation, another CRC researcher approached the first author to say:

After the session, I started to think about metadata, and I started doing it, but I am not sure if I am doing right and how to do it, where the metadata should be stored or how to better organise my data" (Informal meeting with a PhD student, Sociology).

As highlighted in our previous work (Mosconi et al. 2019) the tools that researchers use daily do not offer the possibility to enter metadata and link them to each data item. Metadata writing is a task currently being done, if at all, in the end of the research process shortly before the archival submission. What Data Story suggests is an interface through which uploading a specific data item and engaging with metadata work while still working on the research process is possible and desirable. It is also available for sharing information with colleagues and/or an external audience in a timely fashion.

Lastly, researchers suggested to provide info boxes that could explain in detail the technical features, such as the DOI, the metadata and the tags, so that users could learn about them and understand why they are there and how to make use of them. Other info boxes might be included in the data upload section to explain anonymization, ethical and legal policies. These are important aspects that are often not explained anywhere. They influence how to curate the data and what one can share, but researchers often lack knowledge. In fact, in multiple occasions, researchers asked the project INF to organise seminar sessions on this specific issue which proves the need for more information, training, and support in the field of RDM and the technicalities involved.

4.3 General issues, concerns, and fears

Early on, we decided to provide researchers with a vague definition of what a Data Story actually is in order to allow participants to come up with their own scenarios. However, especially during the thinking aloud evaluation sessions, basic questions came up from the beginning: What is a Data Story? What does it do? Why and how should I write one? For most researchers the three-chapter structure (project set-up, data processing and findings) resonated too much with the structure of academic papers and they wondered in what way a Data Story differs.

Besides stylistic choices, some researchers struggled with the documentation and with the selection of data to show in their Data Stories. For example, Paul asked: *“How would I document that so that people actually understand the interesting insights I had with this story?”*. Paul and a colleague participated at a summer school where they had to illustrate a case study on users’ interactions with apps and present the methodology. They wrote a presentation but, they said, it was hard to convey some of the most interesting questions they had from the dataset, conceptually but also methodologically. During the focus group, the guiding tips were proven helpful in supporting researchers in crafting their narrative and the structuring of the data processing chapter. However, researchers suggested to have a clear separation between the data uploaded and the insights derived from it so that potential reader could better distinguish between a piece of data, personal interpretation, and reconstruction of the analytical process.

To better accommodate Data Stories that are connected to ongoing research, Oliver encouraged us to offer the possibility of starting writing data stories from the data and method section, because:

To what extent do I have to know my story in advance? Am I able to create my story by feeding new bits and pieces and kind of bringing them into an order and swapping them around this storyline until I find that it has somehow become a narrative? That would be something I’d love to know from a design perspective. If it would somehow help to find the narrative, that’s something that could be really interesting as a tool (Thinking aloud session, August 2021, Post Doc in Media History).

In his view, this would potentially allow for bottom-up categories to emerge and to use the Data Story also as an analytical tool. Again, this refers to personal benefits that researchers might see while engaging in data work and their interests in having tools that could support ongoing research.

Our participants also voiced some opinions about Data Story becoming commonplace in academia. Max stressed how some features, similar to those found on social media, could hinder user engagement because some academics might not want to be exposed. Finally, in the focus group, the fear of losing control of the data and data protection came up as an important topic. Concerning this, Max suggested a feature called "visible for a day" because some people might feel uncomfortable "with having data openly accessible in perpetuity".

5 Discussion

The findings illustrated above demonstrate the evolving nature of user reaction to the design as it iterated. As we have stressed, because of our participation as members in the institution, our ongoing interactions with CRC members, and our active research into the issues over a long period of time, we conceive of our efforts as being 'embedded' (Lewis and Russell, 2011). This means that separating evaluation from other investigative processes was neither possible nor desirable. Data Story became both the topic and the medium through which we were able to understand how data curation and sharing practices can be introduced in researchers' daily workflows and how researchers can profit from them. Our contribution highlights lessons learnt through our embedded engagement and provides a new design approach for RDM and for new research data practices. This implies 1) establishing a consensual and gradual process for data curation practices to unfold over time; 2) negotiating metadata readability, flexibility, and standardization through interface design; 3) prompting conversations and learning opportunities with and about data.

5.1 Introducing RDM into collaborative research practices: Lessons Learned

Our initial aim with Data Story was, then, to investigate the priorities that researchers had in respect of data curation, sharing and reuse. These RDM endeavours require the acquisition of data management skills, but the current scientific culture and rewards system do not directly incentivise or yet, recognise these endeavours (Fecher et al., 2017; Feger et al., 2020; Kervin et al., 2014). We had no preconceptions about researchers' priorities but had, in previous work, identified many of the issues they faced when confronting a top-down mandate (Mosconi et al., 2019). We saw the initial scepticism of some researchers but also a recognition of potential benefits connected to sharing and collaboration research practices that are otherwise not traditionally considered in the RDM discourse. In fact, researchers showed interest in learning from others how to do research, how to meaningfully show their own work to others, how best to collaborate together asynchronously, and how to provide an overview of what has been done. The emphasis is also on the user-orientation with transparency in roles and profiles of data workers and collaborators (RfII 2016). The 'data overview' is something that both researchers who collect the data and others interested in the data struggle with. At times, researchers come up with informal practices to visualise their own fieldwork activities and their most important data (as shown by Franziska, which stressed the need for a technical aid like Data Story). As our research participants confirmed, the effort of curation, facilitated and supported through Data Stories, can positively impact how researchers work, and can repay them in providing a structure, assisting them in keeping their data organised or deepen their analysis. In turn, it could make the process of writing publications faster because people can organise and reflect on their findings in and through their curatorial activities elaborated with written narratives.

5.1.1 Curation as consensual and gradual process

Our findings suggest that a solution like Data Story will need firstly to provide features that researchers benefit directly from (i.e.: having the overview, drafting papers, collaborate etc.) and then gradually also introduce curation elements. It also requires a long-term processual perspective for RDM activities which allows researchers to learn new practices as part of their membership of the research infrastructure (Feger et al., 2020; Mosconi et al., 2019). Thus, a gradually emerging consensus around mutual benefit, we anticipate, will consolidate RDM practices and provide learning opportunities (Cox and Verbaan, 2018). The first thinking aloud evaluative sessions focused on a very advanced version of the Data Story concept and the related prototype. It had plenty of metadata. It had a lot of different sections. It had metadata for the story and metadata for files, leading to non-uniformity in practices for metadata curation. Researchers found this type of non-uniformity in data descriptions and the amount of it quite overwhelming. They were confused about the purpose of a Data Story in the first place and wondered why one would put to so much effort into it. Indeed, our earliest prototype proved somewhat paralysing and counter-productive because it attempted to provide an all-encompassing solution. We subsequently adopted what one might term a 'gradualist' solution, one which emphasised the immediate benefits of sharing by focusing on the Data Story as an iterative process, focused on what researchers were interested in but which also, through flexible design, would allow for the addition of other elements. The gradual expansion of metadata is an example of this. With time, from within, we anticipate that we will be able to build a workflow process, based on new standards and new practices of curation and sharing that can be data-driven, negotiate between top-down policies and bottom-up practices, and that can grow and evolve so as to service more distant needs (Pryor 2014; Pryor et al., 2013).

5.1.2 Negotiating metadata readability, flexibility, and standardization

The work of Koesten and Simperl (2021) has previously stressed the importance of narratives and textual documentation needed in order to facilitate data sharing and reuse. Data Story embraces this finding and supports the elaboration of narratives, conceived as "readable metadata for human interpretation", which can highlight the "social function of data" (Birnholtz and Bietz, 2003). Especially with qualitative data, narratives are the vehicle through which researchers perform interpretations, engage reflexively and elaborate data through sense-making (Pepper and Wildy, 2009). The guiding questions (called 'tips') included in the interface design (see section 3.2) aim specifically at supporting such a narrative structure by helping researchers to explicate and organise the implicit knowledge gathered through interactions and observations in the field.

There were evident issues in the emergent logic of the Data Story in relation to, on the one hand, the need for some kind of structure but, on the other, the need for a flexibility in representation which allowed researchers to order matters in ways which were relevant to their work. That flexibility, allowing for their rationales to become visible in their ordering practices, was a useful adjunct in respect of acting as a medium for their own reflections, providing an ongoing, visibly historical document, and providing a medium for engagement with others at various points in project endeavours (Whyte 2014). Specific benefits brought out included the idea that the Data story provided a quick overview, obviating the need for tedious reading; provided a prop for future data collection and analysis; and could replace other forms of sharing which are typically more difficult to find and access. These added degrees of flexibility, however, will need to be negotiated and balanced with some requirements of standardization, for example represented by the metadata elements, which are needed specifically for data retrieval. As suggested by Max, we plan in our future work to identify (through participatory workshops) relevant categories and metadata standards useful to describe methods and data that will be used in conjunction with flexible narratives.

5.1.3 Prompting conversations and opportunities for learning with and about data

As mentioned in section 2, research infrastructures should channel improvements in communication around research artefacts because anything that is shared can in principle be of interest for somebody else so both data creator and recipient need to be allowed to exchange information (Birnholtz and Bietz, 2003). Data Story, even at an early stage, seemed to prompt reflections and conversations about data and its uses. Participants argued that it both stimulated and facilitated conversations with colleagues (and others), encouraged them to be more reflective about their data (the act of building the Story was itself part of an ongoing analytic process), prompting precisely the kinds of thinking about data that methodologies such as grounded theory (see e.g., Muller and Kogan, 2010) seem to recommend. As researchers like Max said, *"it encourages you to think of data, what is the most interesting insights in your data"*. Highlighting what are the most interesting insights from the data at hand is otherwise difficult, especially when drawing the attention of others to it. Data Story encourages researchers to record thinking through practices such as dropping notes into it. It makes data-work visible and present and, as such, facilitates the building of analytic insights while being in conversation (with yourself or) with someone else. We conceptualise these various opportunities as 'reflective' learning opportunities (Boyd and Fales, 1983). Reflective learning is the internal examination and exploration of a concern prompted by an experience, which produces and clarifies meaning in terms of self and leads to a shift in conceptual viewpoint (Boyd and Fales, 1983). In fact, not everyone is equally familiar with the ways in which data is collected, organised, and used in research. In the interdisciplinary contexts we have been involved with, dealing with qualitative data is a new experience for many new researchers and the existence of prior examples which provide rationale for methods adopted or for analytic choices made has proven valuable. Therefore, Data Story can be thought of as an interface which affords learning opportunities (with and about data) of many kinds, above all in relation to research methodology and RDM. It encourages researchers to sit together with their data, curate them and share them, while at the same time supporting them with the organization of their materials and reflection on what they are sharing, who are their sharing with and why. As we move on in this RDM era, data skills are crucial but to learn them, we will need more than just standard routines or pre-defined guidelines, fixed metadata, and categories. As data (and data skills) are the results of ongoing, even serendipitous, learning opportunities and personal (internal) explorations - in relation with a vast ecology of tools, methods, practices - in constant evolution.

6 Conclusion

Solutions to support RDM collaborative workflows are clearly needed. First and foremost, these solutions need to provide benefits to data creators in order to motivate them in using them (Feger et al., 2020). As already highlighted by Rolland and Lee (2013) “investigators need ways to engage in data curation in support of tomorrow’s research without delaying today’s.” (p. 443). In the above, we have demonstrated the opportunities and challenges associated with an alternative approach to RDM which might support these activities in a meaningful way.

Our work is predicated on an investigative policy we have called ‘embedded evaluation’, involving ongoing work by ourselves and others as joint participants to a number of research projects where data curation, sharing and potential reuse has become an issue. Our design was guided by an attempt to negotiate between various interests, and it was in a sense constrained by the funding agency agenda, the INF goals connected to it, and researchers’ concerns and wishes. Our motivation for the work emanated from the realization that the people we worked with in a largely interdisciplinary context are often not trained in, nor used to, data curation and sharing. For the most part they have few resources with which to develop an understanding of the way qualitative data can be organised, what it might be used for, or who it might be used by, nor there are solutions yet that really support the development of a (data) sharing culture within and beyond research groups. What we describe are some steps thus far taken towards meeting that objective. In fact, Data Story offers a simple, and structured way to gain, so to speak, a flavour of the work in question, its epistemic assumptions, its methodologies and specific methods, and its positioning with respect to other work. Naturally, future potential re-users should be kept in mind. We foresee that Data Story can potentially be used for what we would term ‘anticipatory’ articulation work, meaning supporting not only articulation work in respect of current cooperation, but also the work for future cooperation not yet known. The point there is that, in normal organizational life, the kinds of articulation work that are necessary are more predictable. Roles and responsibilities, at least to a degree, are known. That is not the case here. There is no clear agreement about what the responsibilities of active researchers might be, and it is very difficult to anticipate what uses shared data might be put to, and who by. In this sense ‘anticipatory’ articulation work would refer to the work to make future cooperative work possible, in a situation where data work will be fluid, dynamic and mediated by heterogeneous purposes. The Data Story, we argue, provides an entry point into the sensemaking work that will be needed. The focus, then, is on a development from ‘anticipation work’, i.e., “the practices that cultivate and channel expectations of the future, design pathways into those imaginations, and maintain those visions in the face of a dynamic world” (Steinhardt and Jackson 2015, p. 443). We plan in our future work to examine practical implications for research collaboration and RDM in more detail by looking at the kinds of sensemaking that go into narrative structures and the way they are received by others in real contexts.

To conclude, the Data Story, as we call it, is predicated on an amalgam of some orthodox data science constructions and a more flexible, narrative approach. The latter aims to embed the history and the emergent rationale behind the organization of the data and that can highlight “the social function of data in the community that created it” (Birnholtz and Bietz, 2003).We do not imagine that the Data Story will, in and of itself, produce radical and systemic changes to data curation, sharing and reuse practices. Data curation and sharing practices are very much contingent on when and for what reason, and with whom data is to be shared (there will, for instance, be a significant difference between sharing data with other team members, re-using data oneself, and curating it for unknown future users). We do, however, see, in embryo and along with our colleagues, how we can address the need to start developing sharing and RDM strategies step by step, building bottom-up communities of (data) sharing practices in and through the progressive adoption of the solution we describe. We take on board the injunction of Feger et al. (2020) regarding the transition to effective digital RDM and the role of HCI in it: we, as HCI and CSCW researchers, can facilitate the design of interfaces that can support collaborative data work, learning opportunities, encourage reflective thinking, and making data work visible, so that it can be better organised, meaningful, and worthy of our time.

Declarations

Acknowledgements

This research has been possible thanks to the engagement of many scholars who have contributed to shaping this work by showing interests in our approach, and sharing their thoughts through interviews, formal or informal meetings and/or seminars’ interactions. The findings in this paper originate from the project INF funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – (Project ID- 262513311 - SFB 1187). The responsibility for all content supplied lies with the authors.

Ethical Approval

A written informed consent has been used and provided to all participants who have been made aware of the purpose of the study, data analysis procedure, data protection agreements and long-term preservation of the data collection. All participants agreed to take part in the study and consented to the publication of the analysed and anonymized data hereby presented.

Competing interests

The authors declared that they have no conflict of interest.

Authors' contributions

Gaia Mosconi as the lead author conducted all the evaluation activities mentioned in the paper and conceptualized the design of the Data Story. Gaia Mosconi and Aparecido Fabiano Pinatti de Carvalho contributed substantially to all sections of the manuscript. Hussain Abid Syed contributed to the introduction and to the discussion. Gaia Mosconi, Dave Randall, and Helena Karasti engaged in several iterations of data analysis. All authors reviewed the manuscript several times and participated with critical comments and suggestions which greatly improved the paper throughout the process. The order of listing reflects the amount of contribution provided by each author.

Funding

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). Grant ID - 262513311 - SFB 1187.

Availability of data and materials

Not applicable.

References

1. Barry, Danika; Leighann E. Kimble; Bejoy Nambiar; Gareth Parry; Ashish Jha; Vijay Kumar Chattu; M Rashad Massoud; and Don Goldmann (2018). A framework for learning about improvement: embedded implementation and evaluation design to optimize learning. *International Journal for Quality in Health Care*, vol. 30, no. suppl_1, pp. 10–14. https://doi.org/10.1093/intqhc/mzy008
2. Begley, C. Glenn; and Lee M. Ellis (2012). Raise standards for preclinical cancer research. *Nature*, vol. 483, no. 7391, pp. 531–533. https://doi.org/10.1038/483531a
3. Bejarano, Carolina Alonso; Lucia Lopez Juárez; Mirian A. Mijangos García; and Daniel M. Goldstein (2019). *Decolonizing ethnography: Undocumented immigrants and new directions in social science.* Durham: Duke University Press.
4. Birnholtz, Jeremy P.; and Matthew J. Bietz (2003). Data at work: Supporting sharing in science and engineering. In M. Pendergast, K. Schmidt, C. Simone and M. Tremaine (eds): *GROUP'03: Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, Sanibel Island, FL, United States, 9–12 November 2003. New York: ACM Press, pp. 339–348. https://doi.org/10.1145/958160.958215
5. Borgman, Christine L. (2010). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA, USA: MIT Press.
6. Borgman, Christine L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology,* vol. 63, no. 6, pp. 1059–1078. https://doi.org/10.1002/asi.22634

7. Borgman, Christine L. (2015). *Big data, little data, no data. Scholarship in the networked world*. Cambridge, MA, USA: MIT Press.
8. Borgman, Christine L.; Andrea Scharnhorst; and Milena S. Golshan (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, vol. 70, no. 8, pp. 888–904. https://doi.org/10.1002/asi.24172
9. Boyd, Evelyn M.; and Ann W. Fales (1983). Reflective learning: Key to learning from experience. *Journal of Humanistic Psychology*, vol. 23, no. 2, pp. 99–117.
10. Chawinga, Winner Dominic; and Sandy Zinn (2020). Research data management at a public university in Malawi: the role of “three hands”. *Library Management*, vol. 41, nos. 6–7, pp. 467–485. https://doi.org/10.1108/LM-03-2020-0042
11. Chambers, Fred (1994). Removing confusion about formative and summative evaluation: Purpose versus time. *Evaluation and Program Planning*, vol. 17, no. 1, pp. 9–12. https://doi.org/10.1016/0149-7189(94)90017-5
12. Concannon, Shauna; Natasha Rajan; Parthiv Shah; Davy Smith; Marian Ursu; and Jonathan Hook (2020). Brooke Leave Home: Designing a Personalized Film to Support Public Engagement with Open Data. In: *CHI’20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* Honolulu, HI, USA, 25–30 April 2020. New York: ACM Press, pp. 1–14. https://doi.org/10.1145/3313831.3376462
13. Cox, Andrew; and Eddy Verbaan (2018). *Exploring research data management*. London, UK: Facet publishing.
14. Collaboration, Open Science (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, vol. *7*, no. 6, pp. 657–660. https://doi.org/10.1177/1745691612462588
15. Curdt, Constanze; and Dirk Hoffmeister (2015). Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database. *Program: Electronic Library and Information Systems*, vol. 49, no. 4, pp. 494–512. https://doi.org/10.1108/PROG-02-2015-0016
16. Dykes, Brent (2015). Data storytelling: What it is and how it can be used to effectively communicate analysis results. *Applied Marketing Analytics*, vol. 1, no. 4, pp. 299–313.
17. Duarte, Nancy (2019). *Data Story: Explain data and inspire action through story*. Ideapress Publishing.
18. Donner, Eva Katharina (2022). Research data management systems and the organization of universities and research institutes: A systematic literature review. *Journal of Librarianship and Information Science*, vol. 0, no. 0. https://doi.org/10.1177/09610006211070282
19. Dourish, Paul; and Edgar Gómez Cruz (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, vol. 5, no. 2. https://doi.org/10.1177/2053951718784083
20. Eberhard, Igor; and Wolfgang Kraus (2018). Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien. *Mitteilungen der Vereinigung*

Österreichischer Bibliothekarinnen und Bibliothekare, vol. 71, no. 1, pp. 41–52. https://doi.org/10.31263/voebm.v71i1.2018

21. Edwards, Paul N.; Steven J. Jackson; Melissa K. Chalmers; Geoffrey C. Bowker; Christine L. Borgman; David Ribes; Matt Burton; and Calvert Scout (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor: Deep Blue.
22. Edwards, Paul N.; Matthew S. Mayernik; Archer L. Batcheller; Geoffrey C. Bowker; and Christine L. Borgman (2011). Science friction: data, metadata, and collaboration. *Social studies of science,* vol. 41, no. 5, pp. 667–690. https://doi.org/10.1177/0306312711413314
23. Erete, Sheena; Emily Ryou; Geoff Smith; Khristina Marie Fassett; and Sarah Duda (2016). Storytelling with Data: Examining the Use of Data by Non-Profit Organizations. In: *CSCW'16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing,* San Francisco CA, USA, 27 February – 2 March 2016. New York: ACM Press, pp. 1273–1283. https://doi.org/10.1145/2818048.2820068
24. Fecher, Benedikt; Sascha Friesike; Marcel Hebing; and Stephanie Linek (2017). A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications*, vol. 3, no. 1, pp. 1–10. https://doi.org/10.1057/palcomms.2017.51
25. Feger, Sebastian S.; Sünje Dallmeier-Tiessen; Albrecht Schmidt; and Paweł W. Woźniak (2019). Designing for Reproducibility: A Qualitative Study of Challenges and Opportunities in High Energy Physics. In: *CHI'19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland, UK, 4–9 May 2019. New York: ACM Press, pp. 1–14. https://doi.org/10.1145/3290605.3300685
26. 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*, vol. 4, no. CSCW2, pp. 1–26. https://doi.org/10.1145/3415212
27. Feger, Sebastian S.; Paweł W. Woźniak; Jasmin Niess; and Albrecht Schmidt (2021). Tailored Science Badges: Enabling New Forms of Research Interaction. *DIS'21: Designing Interactive Systems Conference 2021,* pp. 576–588. https://doi.org/10.1145/3461778.3462067
28. Fekete, Jean-Daniel (2004). The InfoVis Toolkit. In: *INFO VIS'04: Proceedings of the IEEE Symposium on Information Visualization,* Austin, TX, USA, 10-12 October 2004. New York: IEEE Press, pp. 167–174. https://doi.org/10.1109/INFVIS.2004.64
29. Fekete, Jean-Daniel; Jarke J. van Wijk; John T. Stasko; and Chris North (2008). The value of information visualization. In A. Kerren; J.T. Stasko; JD. Fekete; and C. North (eds): *Information Visualization.* Berlin, Heidelberg: Springer, pp. 1–18.
30. Fortun, Mike; Lindsay Poirier; Alli Morgan; Brian Callahan; and Kim Fortun (2021). What's so Funny 'bout PECE, TAF, and Data Sharing? In D. Boyer; and G. Marcus (eds): *Collaborative Anthropology Today: A Collection of Exceptions.* Ithaca, New York, USA: Cornell University Press.

31. Garza, Kristian; Carole Goble; John Brooke; and Caroline Jay (2015). Framing the community data system interface. In: *British HCI'15: Proceedings of the 2015 British Human Computer Interaction Conference,* Lincoln Lincolnshire, United Kingdom, 13–17 July 2015. New York: ACM press, pp. 269–270.
32. Gebru, Timnit; Jamie Morgenstern; Briana Vecchione; Jennifer Wortman Vaughan; Hanna Wallach; Hal Daumé Iii; Kate Crawford (2021). Datasheets for datasets. *Communications of the ACM*, vol. 64, no. 12, pp. 86–92. https://doi.org/10.1145/3458723
33. Grudin, Jonathan (1988). Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In: *CSCW'88: Proceedings of the 1988 ACM conference on Computer-Supported Cooperative Work, Portland,* Oregon, USA, 26–28 September 1988. New York: ACM Press, pp. 85–93. https://doi.org/10.1145/62266.62273
34. Hamad, Faten; Maha Al-Fadel; and Aman Al-Soub (2021). Awareness of Research Data Management Services at Academic Libraries in Jordan: Roles, Responsibilities and Challenges. *New Review of Academic Librarianship*, vol. 27, no. 1, pp. 76–96. https://doi.org/10.1080/13614533.2019.1691027
35. Irani, Lilly (2010). HCI on the move: methods, culture, values. IN: *CHI EA'10: CHI'10 Extended Abstracts on Human Factors in Computing Systems*, Atlanta, Georgia, USA, 10–15 April 2010. New York: ACM Press, pp. 2939–2942. https://doi.org/10.1145/1753846.1753890
36. Jahnke, Lori M.; and Andrew Asher (2012). The Problem of Data: Data Management and Curation Practices Among University Researchers. CLIR: Council on Library and Information Resources https://www.clir.org/pubs/reports/pub154/problem-of-data/. Accessed 8 November 2022.
37. Kaltenbrunner, Wolfgang (2017). Digital infrastructure for the humanities in Europe and the US: Governing scholarship through coordinated tool development. *Computer Supported Cooperative Work (CSCW)*, vol. 26, no. 3, pp. 275–308. https://doi.org/10.1007/s10606-017-9272-2
38. Kaye, Joseph 'Jofish' (2007). Evaluating experience-focused HCI. *CHI EA'07: CHI'07 Extended Abstracts on Human Factors in Computing Systems*, San Jose, CA, USA, 28 April – 3 May 2007. New York: ACM Press, pp. 1661–1664. https://doi.org/10.1145/1240866.1240877
39. Kervin, Karina; Robert B. Cook; and William K. Michener (2014). The Backstage Work of Data Sharing. In S. Goggins; I. Jahnke; D. W. McDonald; and P. Bjørn (eds): *GROUP'14: Proceedings of the 18th ACM International Conference on Supporting Group Work,* Sanibel Island, Florida, USA, 9 – 12 November 2014. New York: ACM Press, pp. 152–156. https://doi.org/10.1145/2660398.2660406
40. Khan, Nushrat; Mike Thelwall; and Kayvan Kousha (2021). Are data repositories fettered? A survey of current practices, challenges and future technologies. *Online Information Review*, vol. 46, no. 3, pp. 483–502. https://doi.org/10.1108/OIR-04-2021-0204
41. Knaflic, Cole Nussbaumer (2015). *Storytelling with data: A data visualization guide for business professionals*. Hoboken, New Jersey, USA: John Wiley & Sons.
42. Koesten, Laura; Emilia, Kacprzak; Jeni Tennison; and Elena Simperl (2019). Collaborative Practices with Structured Data: Do Tools Support What Users Need? In: *CHI'19: Proceedings of the 2019 CHI*

Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019. New York: ACM Press, pp. 1–14. https://doi.org/10.1145/3290605.3300330

43. Koesten, Laura; and Elena Simperl (2021). UX of data: making data available doesn't make it usable. *Interactions*, vol. 28, no. 2, pp. 97–99. https://doi.org/10.1145/3448888
44. Lallé, Sébastien; and Cristina Conati (2019). The role of user differences in customization: A case study in personalization for infovis-based content. In: *IUI'19: Proceedings of the 24th International Conference on Intelligent User Interfaces,* Marina del Ray, California, 17–20 March 2019. New York: ACM Press, pp. 329–339. https://doi.org/10.1145/3301275.3302283
45. Ledo, David; Steven Houben; Jo Vermeulen; Nicolai Marquardt; Lora Oehlberg; and Saul Greenberg (2018). Evaluation strategies for HCI toolkit research. In: *CHI'18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, 21–26 April 2018. New York: ACM Press, pp. 1–17. https://doi.org/10.1145/3173574.3173610
46. Lee, Jae W.; Jianting Zhang; Ann S. Zimmerman; and Angelo Lucia (2009). DataNet: An emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. *AIChE Journal*, vol. 55, no. 11, pp. 2757-2764. https://doi.org/10.1002/aic.12085
47. Lewis, Susan J.; and Andrew J. Russell (2011). Being embedded: A way forward for ethnographic research. *Ethnography*, vol. 12, no. 3, pp. 398–416. https://doi.org/10.1177/1466138110393786
48. Linne, Monika; and Wolfgang Zenk-Möltgen (2017). Strengthening institutional data management and promoting data sharing in the social and economic sciences. *LIBER Quarterly: The Journal of the Association of European Research Libraries,* vol. 27, no. 1, pp. 58–72.
49. Liu, Shixia; Weiwei Cui; Yingcai Wu; and Mengchen Liu (2014). A survey on information visualization: recent advances and challenges. *Visual Computer*, vol. 30, no. 12, pp. 1373–1393. https://doi.org/10.1007/s00371-013-0892-3
50. MacDonald, Craig M.; Michael E. Atwood (2013). Changing perspectives on evaluation in HCI: past, present, and future. *CHI EA'13: CHI'13 Extended Abstracts on Human Factors in Computing Systems*, Paris, France, 27 April – 2 May 2013. New York: ACM Press, pp. 1969–1978. https://doi.org/10.1145/2468356.2468714
51. Mackay, Wendy E.; Caroline Appert; Michel Beaudouin-Lafon; Olivier Chapuis; Yangzhou Du; Jean-Daniel Fekete; and Yves Guiard. (2007). Touchstone: exploratory design of experiments. In: *CHI'07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* San Jose, California, USA, 28 April – 3 May 2007. New York: ACM Press, pp. 1425–1434. https://doi.org/10.1145/1240624.1240840
52. Martinez-Maldonado, Roberto; Vanessa Echeverria; Gloria Fernandez Nieto; and Simon Buckingham Shum (2020). From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. In: *CHI'20: Proceedings of the 2020 CHI Conference on Human Factors in Computing* Systems, Honolulu, HI, USA, 25–30 April 2020. New York: ACM Press, pp 1–15. https://doi.org/10.1145/3313831.3376148

53. Méndez, Gonzalo Gabriel; Uta Hinrichs; and Miguel A. Nacenta (2017). Bottom-up vs. Top-down: Trade-offs in efficiency, understanding, freedom and creativity with infovis tools. In: *CHI'17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, 6–11 May 2017. New York: ACM Press, pp. 841–852. https://doi.org/10.1145/3025453.3025942
54. Moreau, Katherine A.; Kaylee Eady; Lindsey Sikora; and Tanya Horsley (2018). Digital storytelling in health professions education: a systematic review. *BMC Medical Education*, vol. 18, no. 1, pp. 1–9. https://doi.org/10.1186/s12909-018-1320-1
55. 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 (CSCW)*, vol. 28, no. 3–4, pp. 749–789. https://doi.org/10.1007/s10606-019-09354-z
56. Mosconi, Gaia; Dave Randall; Helena Karasti; Saja Aljuneidi; Tong Yu; Peter Tolmie; and Volkmar Pipek (2022). Designing a Data Story: A Storytelling Approach to Curation and Sharing in Support of Ethnographically-driven Research. *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, Article 289 (November 2022).
57. Muller Michael J.; and Sandra Kogan (2010). Grounded Theory Method in Human Computer Interaction and Computer-Supported Cooperative Work. In J. A. Jacko (ed): *Human Computer Interaction Handbook.* Boca Raton: CRC Press, pp. 1003–1024.
58. Pantazos, Kostas; and Søren Lauesen (2012). Constructing visualizations with InfoVis tools: An evaluation from a user perspective. In: *GRAPP 2012 IVAPP 2012: Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*, Rome, Italy, 24–26 February 2012. pp. 731–736.
59. Pasquetto, Irene V.; Ashley E. Sands; and Christine L. Borgman (2015). Exploring openness in data and science: What is "open," to whom, when, and why? *Proceedings of the Association for Information Science and Technology,* vol. 52, no. 1, pp. 1–2. https://doi.org/10.1002/pra2.2015.1450520100141
60. Pepper, Coral; and Helen Wildy (2009). Using narratives as a research strategy. *Qualitative Research Journal*, vol. 9, no. 2, pp. 18–26. https://doi.org/10.3316/QRJ0902018
61. Pinfield, Stephen; Andrew M. Cox; and Jen Smith (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLoS ONE*, vol. 9, no. 12, e114734. https://doi.org/10.1371/journal.pone.0114734
62. Poirier, Lindsay (2017). Devious design: Digital infrastructure challenges for experimental ethnography. *Design Issues*, vol. 33, no. 2, pp. 70–83. https://doi.org/10.1162/DESI_a_00440
63. Preuss, Nils; Georg Staudter; Moritz Weber; Reiner Anderl; and Peter F. Pelz (2018). Methods and technologies for research-and metadata management in collaborative experimental research. *Applied Mechanics and Materials*, vol. 885, pp. 170–183. https://doi.org/10.4028/www.scientific.net/AMM.885.170

64. Pryor, Graham (2014). Who's doing data? A spectrum of roles, responsibilities, and competencies. In G. Pryor; S. Jones; A. Whyte (eds): *Delivering Research Data Management Services: Fundamentals of Good Practice*. London, UK: Facet Publishing, pp. 41–58.

65. Pryor, Graham; Sarah Jones; and Angus Whyte (2013). Options and approaches to RDM service provision. In G. Pryor; S. Jones; A. Whyte (eds): *Delivering Research Data Management Services: Fundamentals of Good Practice*. London, UK: Facet Publishing, pp. 21–40.

66. Remy, Christian; Oliver Bates; Alan Dix; Vanessa Thomas; Mike Hazas; Adrian Friday; and Elaine M. Huang (2018). Evaluation beyond usability: Validating sustainable HCI research. In: *CHI'18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, 21–26 April, 2018. New York: ACM Press, pp. 1–14. https://doi.org/10.1145/3173574.3173790

67. RfII. (2016). *German Council for Scientific Information Infrastructures: Enhancing Research Data Management: Performance through Diversity. Recommendations regarding structures, processes, and financing for research data management in Germany*. Göttingen. https://d-nb.info/1121685978/34. Accessed 21 September 2022.

68. Rolland, Betsy; and Charlotte P. Lee (2013). Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. In A. Bruckman, S. Counts, C. Lampe and L. Terveen (eds): *CSCW'13: Proceedings of the 2013 conference on Computer supported cooperative work,* San Antonio, Texas, 23– 27 February 2013. New York: ACM Press, pp. 435–444. https://doi.org/10.1145/2441776.2441826

69. Rowhani-Farid, Anisa; Michelle Allen; and Adrian G. Barnett (2017). What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review*, vol. 2, no. 1, pp. 1–10. https://doi.org/10.1186/s41073-017-0028-9

70. Ryen, Anne (2011). Ethics and qualitative research. In D. Silverman (ed): *Qualitative Research*, 3rd ed. London, UK; Thousand Oaks, CA, USA; New Delhi; Singapore: SAGE Publications, pp. 416–238.

71. Steinhardt, Stephanie B.; and Steven J. Jackson (2015). Anticipation Work: Cultivating Vision in Collective Practice. In: *CSCW'15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, Vancouver, BC, Canada, 14–18 March. New York: ACM Press, pp. 443–453. https://doi.org/10.1145/2675133.2675298

72. Strauss, Anselm L.; and Juliet M. Corbin (1998). *Basics of qualitative research. Techniques and procedures for developing grounded theory.* Thousand Oaks: SAGE Publications.

73. Thomas, David R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, vol. 27, no. 2, pp. 237–246. https://doi.org/10.1177/1098214005283748

74. Treloar, Andrew; and Cathrine Harboe-Ree (2008). Data management and the curation continuum: how the Monash experience is informing repository relationships. In: *VALA 2008: The 14th Biennial Conference and Exhibition,* Melbourne, 5 – 7 February 2008.

75. 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,* vol. 169, no. November 2016, pp. 191–198. https://doi.org/10.1016/j.socscimed.2016.08.004
76. Twidale, Michael; David Randall; and Richard Bentley. (1994). Situated evaluation for cooperative systems. In: *CSCW'94: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work,* Chapel Hill, North Carolina, USA, 22–26 October 1994. New York, NY, USA: ACM Press, pp. 441–452.
77. Velden, Theresa (2013). Explaining field differences in openness and sharing in scientific communities. In A. Bruckman; S. Counts; C. Lampe; and L. Terveen (eds): *CSCW'13: Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, San Antonio, Texas, 23–27 February. New York, USA: ACM Press, pp. 445–458.
78. Van Den Haak, Maaike; Menno De Jong; and Peter Jan Schellens (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, vol. 22, no. 5, pp. 339–351. https://doi.org/10.1080/0044929031000
79. Wallis, Jillian C.; 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,* vol. 8, no. 7, e67332. https://doi.org/10.1371/journal.pone.0067332
80. West, Christina H.; Kendra L. Rieger; Amanda Kenny; Rishma Chooniedass; Kim M. Mitchell; Andrea Winther Klippenstein; Amie-Rae Zaborniak; Lisa Demczuk; and Shannon D. Scott. (2022). Digital Storytelling as a Method in Health Research: A Systematic Review. *International Journal of Qualitative Methods,* vol. 21, pp. 1–25. https://doi.org/10.1177/16094069221111118
81. Whyte, Angus; Jonathan Tedds (2011). Making the Case for Research Data Management. *DCC Briefing Papers*, Edinburgh: Digital Curation Centre (DCC). https://www.dcc.ac.uk/guidance/briefing-papers/making-case-rdm. Accessed 15 July 2022.
82. Wilkinson, Mark D.; 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, andBarend Mons (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, vol. 3, no. 1, 1–9. https://doi.org/10.1038/sdata.2016.18
83. Wulf, Volker; Claudia Müller; Volkmar Pipek; David Randall; Markus Rohde; and Gunnar Stevens (2015). Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design

Case Studies. In V. Wulf; K. Schmidt; and D. Randall (eds): *Designing Socially Embedded Technologies in the Real-World*. London, UK: Springer London, pp. 111–150.

84. Wu, Jing; and Der-Thanq Victor Chen (2020). A systematic review of educational digital storytelling. *Computers and Education*, vol. 147, April 2020, 103786. https://doi.org/10.1016/j.compedu.2019.103786
85. Xu, Xian; Leni Yang; David Yip; Mingming Fan; Zheng Wei; and Huamin Qu. (2022). From "Wow" to "Why": Guidelines for Creating the Opening of a Data Video with Cinematic Styles. In S. Barbosa; C. Lampe; C. Appert; D. A. Shamma, S. Drucker; J. Williamson; K. Yatani (eds): *CHI'22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems,* New Orleans, LA, USA, 29 April – 5 May 2022. New York, USA: ACM Press, pp. 1–20. https://doi.org/10.1145/3491102.3501896
86. Zimmerman, Ann (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries,* vol. 7, nos. 1–2, pp. 5–16. https://doi.org/10.1007/s00799-007-0015-8
87. Zuiderwijk, Anneke; and Helen Spiers (2019). Sharing and re-using open data: A case study of motivations in astrophysics, International Journal of Information Management. *International Journal of Information Management*, vol. 49, December 2019, pp. 228–241. https://doi.org/10.1016/j.ijinfomgt.2019.05.024

Figures

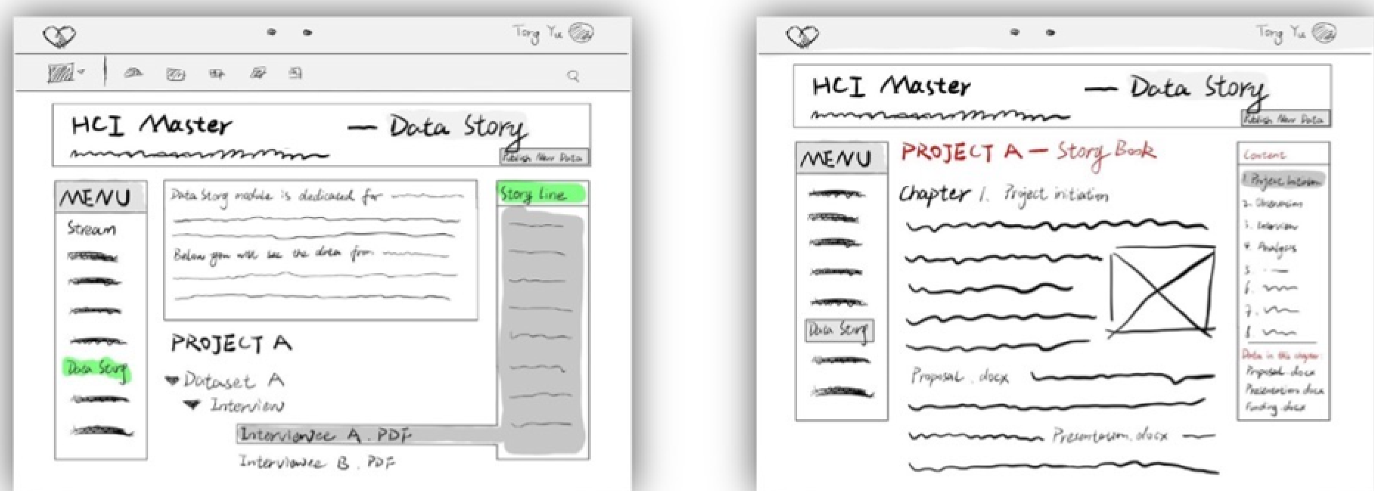

Figure 1

First sketches of the Data Story concept (February 2021).

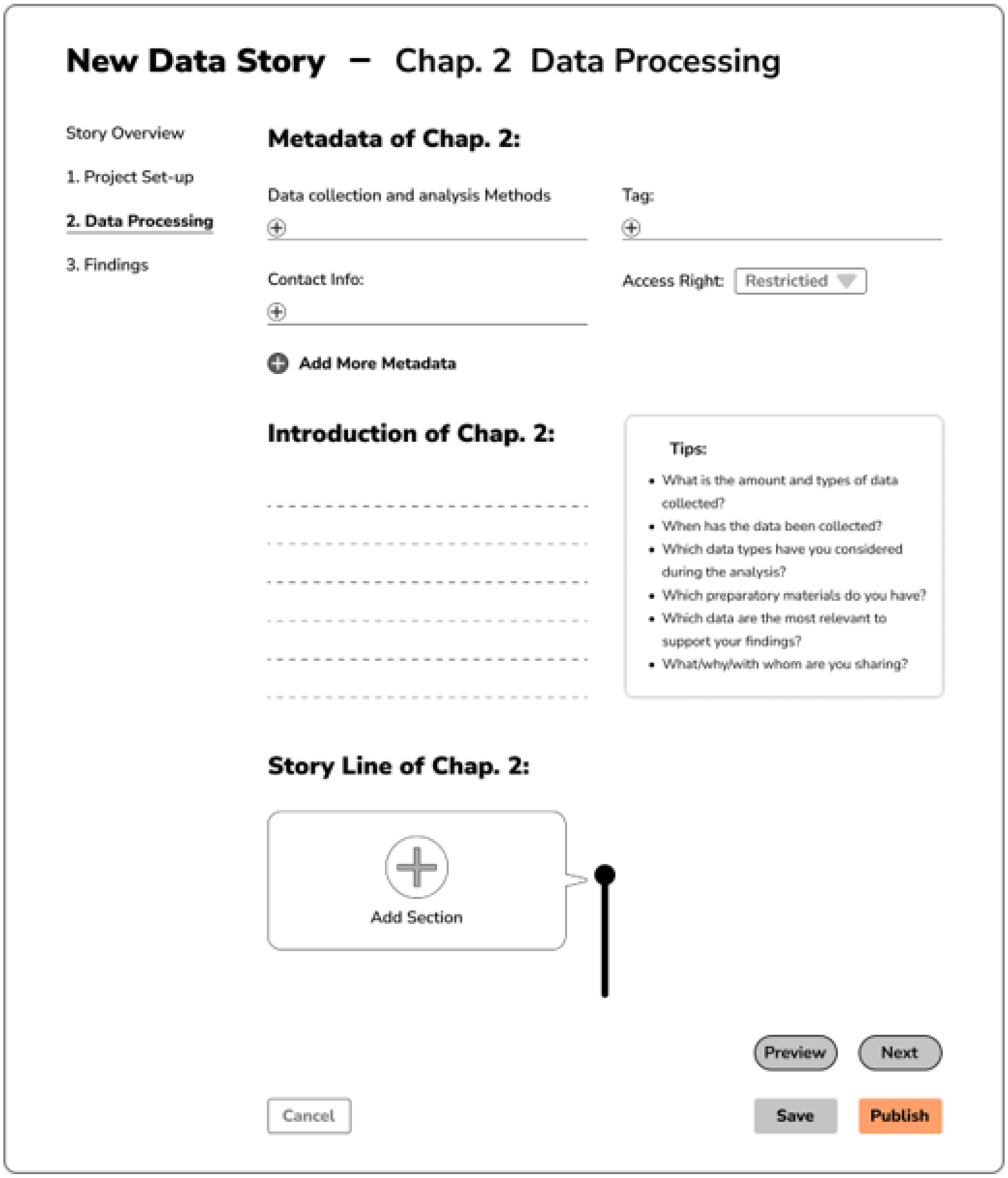

Figure 2

Data Story processing chapter

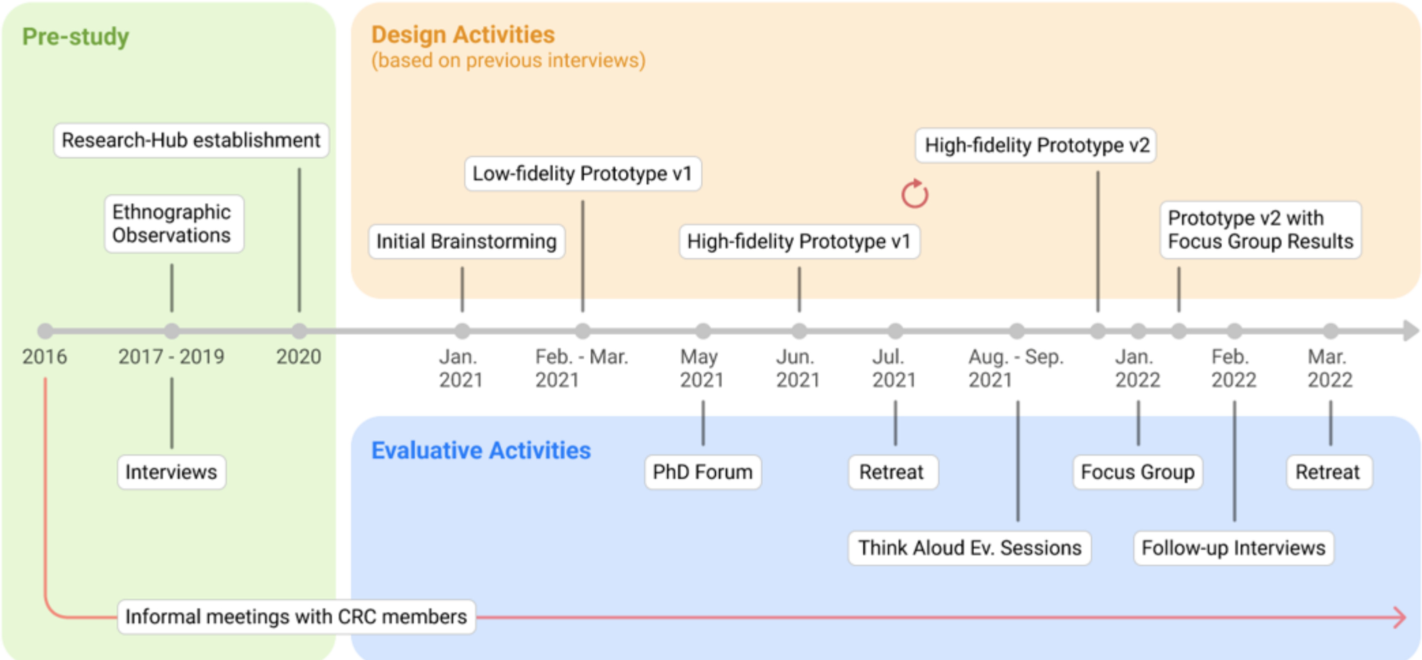

Figure 3

Embedded evaluation: overview of fieldwork, design, and evaluation activities

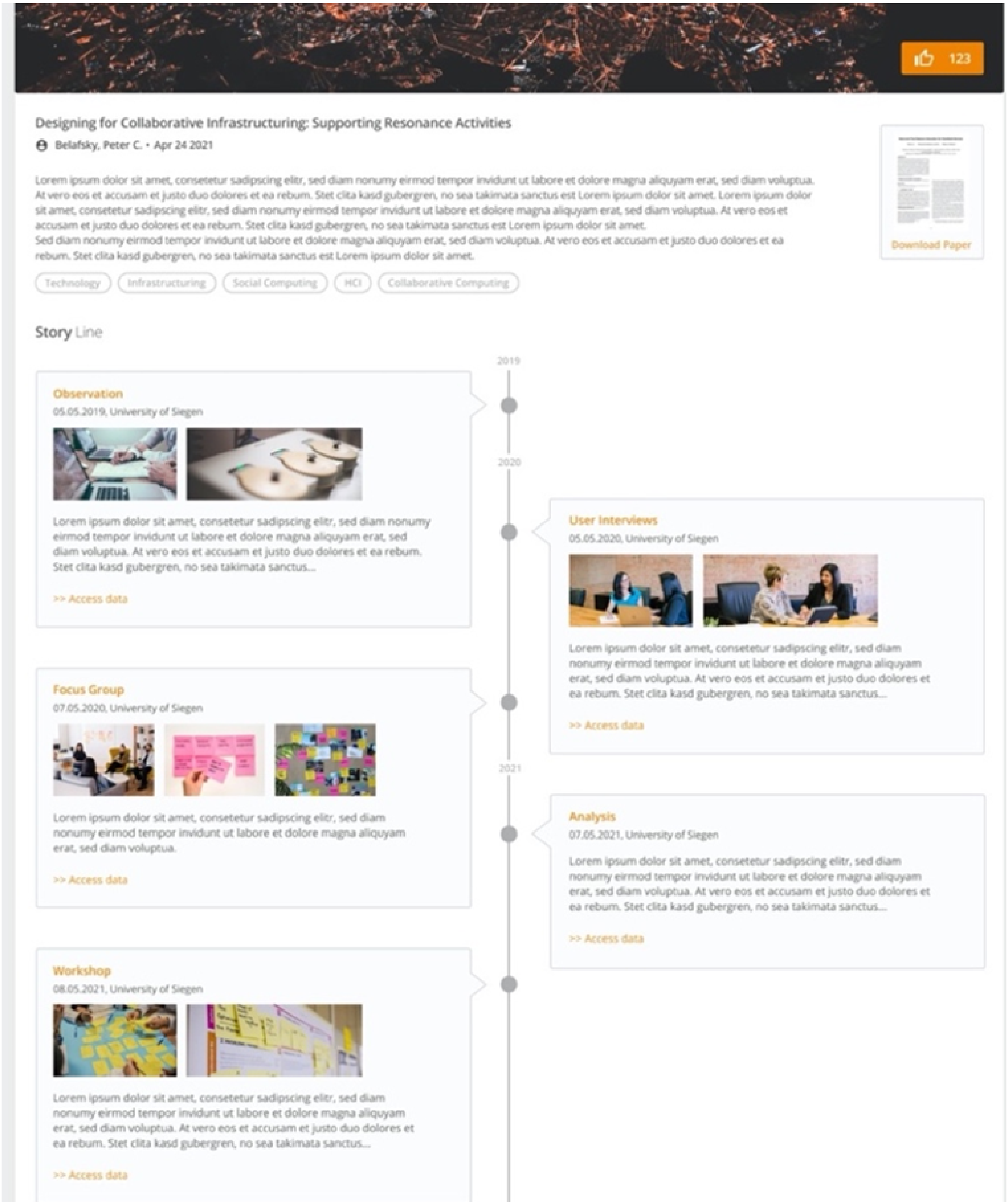

Figure 4

Second version of the high-fidelity prototype redesigned according to the feedback of Thinking Aloud Evaluation Sessions