Tensions in Data Journey Activities: Mobilising, Processing, Producing, and Re-purposing Data in Environmental Assessment Practice

RIKKE HAGENSBY JENSEN, Department of Computer Science, Aalborg University, Denmark
STINE S. JOHANSEN, Queensland University of Technology, Australia
NICOLAI BRODERSEN HANSEN, Department of Computer Science, Aalborg University, Denmark
ASHNA MAHMOOD ZADA, Department of Computer Science, Aalborg University, Denmark
PETER AXEL NIELSEN, Department of Computer Science, Aalborg University, Denmark

363

Digital data shape work practices as they travel through space, time, and social situations. In turn, workers shape the materiality of data through different data-oriented activities. In this paper, we draw on theory of data journeys to develop and present a conceptual framework of four data-intensive activities: Producing, mobilising, processing, and re-purposing. Together, those activities capture entanglements of data transformations and data-intensive practices. We develop the framework through a qualitative study of digitisation in Environmental Assessment (EA) practice. EA practitioners reflect on their experiences working with digital environmental data through ten semi-structured and two interactive interviews. The paper offers two main contributions. First, we develop a data journey theoretical framework and show that it is useful as a unit of analysis for uncovering tensions in data-oriented activities. Second, we use these insights to outline a set of tensions for future design activities in the area of data work practices.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**; *Empirical studies in HCI*.

Additional Key Words and Phrases: Data, data journey, practice, tensions, theoretical framework, environmental assessment

ACM Reference Format:
Rikke Hagensby Jensen, Stine S. Johansen, Nicolai Brodersen Hansen, Ashna Mahmood Zada, and Peter Axel Nielsen. 2023. Tensions in Data Journey Activities: Mobilising, Processing, Producing, and Re-purposing Data in Environmental Assessment Practice. *Proc. ACM Hum.-Comput. Interact.* 7, PRE-PRINT Version accepted for PACMHCI - CSCW2 , Article 363 (October 2023), 22 pages. https://doi.org/10.1145/3610212

1 INTRODUCTION

Digital data and the activities through which they are produced, mobilised, processed, and re-purposed increasingly pervade work practices [10]. To capture that data travel in such practices, previous research has recently coined the term "data journeys" [1, 34]. The term "journey" connotes that data rarely travel in a continuous, linear fashion but rather with pauses, several start and end points, and with friction along the way [14]. As data travel through different situations and between

Authors' addresses: Rikke Hagensby Jensen, Department of Computer Science, Aalborg University, Denmark, rjens@cs.aau.dk; Stine S. Johansen, Queensland University of Technology, Australia, stine.johansen@qut.edu.au; Nicolai Brodersen Hansen, Department of Computer Science, Aalborg University, Denmark, nbha@cs.aau.dk; Ashna Mahmood Zada, Department of Computer Science, Aalborg University, Denmark, amza@cs.aau.dk; Peter Axel Nielsen, Department of Computer Science, Aalborg University, Denmark, pan@cs.aau.dk.

2573-0142/2023/10-ART363 $15.00
https://doi.org/10.1145/3610212

different people, they change material form and, in turn, shape data work practices [10, 45]. Though a large body of work has addressed different aspects of data work practices [5, 39, 41, 47, 59, 75], we argue in this paper that there is still a need for further solidifying a common understanding of the dynamic entanglements and frictions between data and data work activities as they travel.

In this paper, we present a data journey framework which can be utilised as an analytical tool to uncover tensions in data work practices. The framework is an extension of existing research on data journeys, particularly the research by Leonelli [34]. We extend Leonelli's theory through a case study of data work in Environmental Assessment (EA) practices. EA practices rely on a wide variety of data work activities [38] to assess the environmental impact of developmental plans, policies, and construction projects prior to approval. EA reports document the impact of a proposed initiative, including potential risks and mitigation strategies [36], and provide input to decision-making among government bodies, consulting agencies, and the public [9]. In this practice, both quantitative and qualitative data are routinely used to outline the environmental assessment. As EA relies heavily on the production, processing, mobilisation, and re-purposing of data across various stakeholder groups, they are highly conducive to our investigation of data journey activities.

Based on 10 interviews and two follow-up sessions with 14 stakeholders, including consultants, government administrators, developers, and NGO representatives, we gained insights into entanglements between data transformations and data work activities in EA practices. Taking a point of departure in Leonelli's definition of data journeys as *"the movement of data from their production site to many other sites in which they are processed, mobilised and re-purposed"* [34], we conducted a deductive, reflective analysis and identified four data journey activities of producing, processing, mobilising, and re-purposing. We observed how friction in data work emerged and manifested as data tensions for each of the four activities. Based on our investigation, this paper contributes with a solidified theoretical framework, anchored in a data-intensive work practice and validated through subsequent interactive interviews.

By offering a coherent framework for analysing and unpacking tensions in data journey activities, our paper answers recent calls for the CSCW and HCI research communities to take up the matter of data work practices in "*the hopes of developing a taxonomy of work practices and open issues in the behavioural and social study of data science and data science workers*" [40]. Our reasoning for engaging in this endeavour is two-fold. First, a clear definition better enables practitioners outside the research field to articulate how these activities emerge in their daily data work practices. This is helpful in gaining a better understanding of data journeys in particular domains which require expert knowledge. In this paper, we show that this understanding can be expressed as tensions. Second, we argue that a unified understanding is key to evolving the research field. Unclear definitions muddle scientific collaboration and transfer of knowledge. In order to create a unified understanding, one step is to develop it through empirical investigations wherein the ideals and potential pitfalls of the theory can be uncovered. We further discuss how accounting for tensions in data journeys can bring data to the foreground [35, 61, 68] and thus expose new design opportunities for supporting data work.

2 RELATED WORK

Both previous HCI and CSCW research cover different data work routines, including perspectives where data act as a core material in work practices (*e.g.,* data science [41, 46], data analytics [26, 48], and scientific research [5, 34, 43]), and more domain-specific practices where working with digital data have transfused daily work routines (*e.g.,* health domains [23, 39] and environmental assessments [11, 38]). To serve our purpose of conducting an empirical study of a data-intensive work practice, we focus on prior research that explores ways of supporting data work activities, collaborative aspects and the emergence of tensions related to that work, and research that investigates the role data play in shaping work and design practices.

2.1 Supporting Data-Oriented Work

One area of research investigates how the design of new digital technologies can support data workers in their daily routines. For instance, Kandel et al. [26] describe the interactive tool 'Wrangler' that assists data analysts in cleansing and transforming data when addressing data quality issues. Wang et al. [71] introduce 'AutoDS', a tool that uses ML automation to support and improve data scientists' data-oriented work activities. Both studies illustrate that digitised tools designed to support data workers, shape how data transform in specialised data-oriented activities. Other research illustrates how domain-specific workers make use of digital tools to foster data work in a specific domain. For instance, Sanches and Brown [57] studied how the digitalisation of malaria data is operationalised as disease surveillance in diverse human, organisational, and infrastructural situations through dedicated care. Similarly, Ismail and Kumar [23] illustrate the demands and conflicts front-line healthcare workers in Delhi (India) experience in their data collection practices. Related to EA practice, Mehra et al. [38] show that augmented and mixed reality technologies are useful for creating interactive data representations to encourage more collaborative, time-efficient, and transparent assessments. Similarly, Law et al. [32] use augmented reality to support EA practitioners in interpreting and comparing collected environmental data. This prior research of CSCW perspectives and HCI evaluations demonstrates that data workers' interactive tools support how data work is performed for different purposes and in different contexts. However, as these supportive technologies are finding their way into many aspects of daily work routines, most endeavours do not fully engage with broader framings to understand the social and dynamic aspects of data transformation and data work as data travels between activities.

A number of papers present investigations that explore how social dynamics unfold in diverse teams that work jointly on data-intensive projects. They do so by primarily focusing on understanding the creative act of situated collaborations in data work practices, and how digitised tools support these collaborations [6, 37, 43, 59, 64, 75]. For instance, Zhang et al. [75] surveyed 183 participants working with different aspects of a data science project and found that teams are exceptionally collaborative in such projects as they utilise a variety of tools to support these aspects of their data work practices. Neff et al. [43] studied academic data scientists and cross-disciplinary engineering teams. The authors emphasise different complexities of contextualised data-oriented work that shape cultures around these disciplines. They suggest a set of concepts for analysing data science practices (*e.g.*, sense-making of data as a collective process, and viewing data as stories). Taylor et al. [64] describe the relationship between communities and data, illustrating how data materialise in distinct ways dependent on situations and are shaped through situational, physical and digital components. Looking towards data-oriented work related to environmental assessment, Dourish et al. [11] explore the collaborative nature of EA practices when sharing and appropriating documents for large-scale engineering projects. Together, these studies show the complexity of data changing material form in the hands of different collaborators. Further, such changes are shaped by diverse work contexts and collaborators' specialised competencies also forming these practices.

Focusing on data as material, previous HCI research has also framed data as a central material for future design to support data workers [7, 17, 20, 27, 30, 61]. Making data an active element in the design process, Kun et al. [30] describe their design framework 'Exploratory Data Inquiry', which supports novice designers' creativity when designing for future data-intensive work. Similarly, Seidelin et al. [61] explore foregrounding data as part of collaborative service design, suggesting that when data become an explicit design element, domain experts are more inclined to view data as active material in a co-design process. Likewise, Khovanskaya and Sengers [27] illustrate how data rhetoric of data-driven practices of mid-century American labour unionists calls for opportunities for participatory design to promote workplace advocacy. To explore alternative visions for designing with data, Lupton et al. [35] explore more-than-human perspectives as an alternative way of re-imaging digital data. Similarly, Feinberg [17] shows that data, when viewed from a design perspective, emerges as part of interwoven activities, narrating the story that data are indeed dynamic and situational in essence.

The breadth of the research outlined above illustrates the already vast interest in investigating data-intensive work practices. Collectively, this prior research demonstrates that when data become an embedded element in a diverse set of daily work routines, data are shaped by the situations, tools, disciplinary rites, and competencies embedded in the situated practices. However, most of these studies do not address how data change material form as they travel between data work activities, including the friction they encounter. Further, the above studies illustrate that it is not unproblematic to move data between activities. We believe that the conceptualisation of a coherent framework that accounts for how data journey can bring alternative insights into the dynamic entanglements of data work and frictions, as data move between activities. To better account for how data travel can be understood as social and material elements simultaneously, we are inspired by recent work by Seidelin et al. [61], Lupton et al. [35], Neff et al. [43], and Feinberg [17] arguing for foregrounding data in our research. Nevertheless, we believe there is still a need to extend this research by developing approaches that allow researchers and practitioners to better articulate the dynamic roles of data in their daily work. Such approaches would allow for more participatory engagement between researchers, designers, and practitioners.

2.2 Transformations of Data and Work Practices

Another important body of prior research concerns the digitisation of data, transforming how people work with data in many companies and institutions. Such digital transformations have been framed in prior research as a technology-driven, continuous process that triggers significant changes in all areas of human society [55, 63, 69]. This suggests that digital transformation is not only organisation-centric (*e.g.,* changing organisational processes to drive better operational performance [13]), but also social (*e.g.,* ensuring a more innovative and collaborative culture [13, 69]) and technological (*e.g.*, using innovative digital technologies like big data, analytics, cloud computing, etc. [55]).

To capture socio-technical aspects of data work practices in organisations in HCI and CSCW, recent research is concerned with the ways data-oriented workers engage with data as part of their work routines with the aim of building knowledge about these work practices. Muller et al. [41] conducted interviews with data scientists to conceptualise five human interventions (discovery, capture, curation, design, and creation) to frame data as a human-influenced entity. The authors argue that each of these human interventions requires different expertise, tools, and craftsmanship depending on the degree to which data science practitioners need to cleanse their data for a given project. Passi and Jackson [47] reveal that trust in collaborative data science practices is reliant not only on data scientists' analytical skills to pre-process and quantify data but also on their ability to negate and translate data into shared understandings with other stakeholders. Sambasivan et al. [56] coin the term 'Data Cascades' as compounding data deteriorating events activated by AI practices that undervalue data. Related to this, Muller and Strohmayer [42] frame 'forgetting' in data science practices to understand how bias in data sets also emerges from human-centred activities, highlighting that data science work is a socio-technical practice. Studying data-oriented work in hospitals, Møller et al. [39] argue that domain practitioners play a major role in explaining where data come from and ensuring that data-based conclusions are sensible and reasonable. Likewise, Seidelin et al. [60] demonstrate how data practices in the public sector play a vital role in how political negotiations and decisions come about.

In addition to the above, some papers describe the friction and tensions which shapes data and work practices. Pink et al. [52] demonstrate the concept of "broken data" frictions as data increasingly mobilise in ways that impact society and people's everyday life by *"accounting for how data might be in processes of decay, making, repair, re-making and growth"*. Similarly, Edwards et al. [15] argue that data tensions and conflicts embedded in data work routines of science practices are often resolved at the data level as data transform into things (*e.g.*, samples, specimens, and collections), and the outcome materialised as visual, tactile, numeric, statistical models or "higher-order" products (*e.g.*, patents and even publications). This illustrates that as data tensions emerge and are acted upon, different meanings also shape how data are transformed and interpreted. For this reason, we believe further research would benefit greatly from enabling practitioners to articulate and cultivate tensions as data move between activities.

To address the different perspectives for understanding socio-material elements of work practices in our study, we draw on research that articulates how practices are entangled, dependent on the situation, and formed through socio-material actions [45]. As a way to address Bødker [3]'s third-wave HCI challenges, a practice-oriented design perspective has recently been advocated in HCI [31, 50] as a complementary theoretical lens to more holistic embrace the relationship of humans, digital materials, and patterns of use. Theories to understand (social) practice originate from social science, where researchers argue that making people's everyday practices a unit of analysis, as opposed to individual users and user needs, leads to knowledge about patterns of routine from socially shared activities [44, 45, 54, 62]. Orlikowski [44] argues that investigating technology use in organisations through a practice-based lens enables a deeper understanding of "*...the role of social practices in the ongoing use and change of technologies in the workplace*", including ways of learning, tool adoptions, changes to digital data, and transformations of the status quo. Shove et al. [62] frame changes in (social) practice as re-configurations of three interconnected elements, *i.e.*, competencies (*e.g.*, shared know-how and skills needed to perform a practice), materials (*e.g.*, objects, technologies, infrastructures, and digital data necessary to perform the practice), and meanings (*i.e.*, socially shared understandings of what is the proper or socially acceptable ways to perform a practice).

We take inspiration from the above research, arguing that data work practice is indeed a dynamic endeavour shaping not only the social and cultural dimensions of data routines but also data's "raw" interactive materiality [10, 19, 21, 40, 72, 73]. Further, as tensions emerge, data transform and create new meanings [51], illustrating that data and work practices are in constant change [15, 52]. To embrace such dynamics in this paper, we situate "data work" as a socio-material practice meaning that we consider data work as the *"constitutive entanglement of the social and the material in everyday organisational life"* [45]. Our understanding of data work practice is thus a broad understanding of different practitioners working with diverse assortments of data in their daily work routines. Further, we recognise that data work practices are diverse, requiring different competencies and materials, and performed in many distinct contexts [5]. Nevertheless, we believe that by framing data work as a social-material practice [45], we embrace the plurality of how these different practices play out in everyday work life.

To better understand the dynamics of data and how they journey as part of practice, transform and possibly meet friction on their way, we turn to how we can frame data as they travel – as a central and yet reciprocal, dynamic element of data work practices.

3 CONCEPTUALISING DATA JOURNEY ACTIVITIES

To set the scene for our investigation, we include this section to outline the concept of data journeys as an integral part of data-intensive work activities. A small body of HCI research showcases the concept of journey mapping to capture the dynamics of data through interactions between humans and digital technology. For instance, Prieto-Alvarez et al. [53] propose a collaborative mapping tool, "Learner-Data Journey", that illustrates the potentials of journey mapping in an educational context. The authors ground their conceptualisation of "Learner-Data Journey" on user journey mapping and report that creating journey visualisations to facilitate communication and involving other stakeholders in a design process leads to valuable knowledge. Also within a learning context, Koningsbruggen et al. [68] devised a teaching method they term "data diaries". They argue that data diaries can be a valuable vehicle to move from visualised to physicalised data materiality, and better allow for focusing on the story of data. In a study of IoT data with householders, Desjardins et al. [8] expose "*the plural, situated, and messy ways data are entangled in and with a home, its inhabitants, and outside actors such as companies and other data*". The authors argue that critical data investigations may bring on plural perspectives as to how "*data are produced, where they travel, and who controls, handles, and manages them*".

Outside HCI, Bates et al. [1] coin the concept of "data journeys" in an effort to better embrace the "life of data". They argue that their data journey conceptualisation contributes to the development of critical, qualitative methodologies involving design-based mapping, oral data histories, and digital ethnography. They propose that such methods can expose the dynamics of data practices to highlight data movements as data "*places and connections are dynamic, evolving over time, and emergent socio-material conditions can open up new possibilities for data journeys through space and time.*" Building on this prior research, Leonelli [34] offers a conceptualisation of data journeys which we utilise to frame data journey activities in this paper. Leonelli argues that making data the unit of analysis for data-oriented activities will lead to an identification of the relationship between the physical and social characteristics of data objects and their movements. These movements are described as "*data journeys, designating movement of data from their* ***production*** *site to many other sites in which they are* ***processed***, ***mobilised*** *and* ***re-purposed***" (our emphasis).

Our aim is to extend the above definition that includes four data journey activities: Producing, processing, mobilising, and re-purposing. These activities are provisionally described by Leonelli as follows. Producing corresponds to where and how data is generated, including conceptual tools (such as theories and methods) and material tools (for measuring or experimenting) [34, p. 16]. Mobilising is when data are transformed for a specific purpose or use case [34, p. 6]. Processing is concerned with how data are computationally interpreted and formatted (*e.g.*, aggregated, computed, or visualised) [34, p. 14]. Finally, re-purposing is the re-use of assembled data [34, p. 5].

Despite numerous HCI studies investigating the digitalisation of data in shaping everyday and work life [3, 8, 15, 27, 39, 41, 44, 61, 75], there has, to our knowledge, not been any investigations that make data journey activities the unit of analysis by focusing on data, how they travel, transform, and create tensions as part of data-intensive activities. We propose that viewing data journeys through the lens of these four data journey activities better captures how data travel across space, time and social situations in data work practices. It is not clear yet, however, how the activities can be used as an analytical tool to do so, in part due to the vague descriptions of what constitutes each type of activity as well as a lack of empirical investigations to support the description of each. Therefore, we aim to further clarify the four data journey activities and how data tensions may emerge as part of these activities through an empirical case. We propose a potential for using Leonelli's four provisionally described data activities as a starting point to analytically understand socio-material elements of data journeys as part of identifying tensions in data work practices.

4 METHOD

4.1 Data Journeys in Environmental Assessment Practice

To serve our goal of extending the concept of data journeys and use that as an analytical tool, we anchor the investigation in the domain of environmental assessment. There are three main reasons why this domain is suitable for our investigation. First, the practice is inherently data-intensive, domain-specific, with multiple types of data to support decision-making. Second, many stakeholders are involved in the process with multiple entry and exit points. Third, EA practice resembles scientific practice [15, 34] and data science work [41, 43, 47, 75] where data and data transformations play a crucial role in forming well-structured arguments.

EA reports are produced to support decision-making when developers propose projects that can affect the environment. Typically, the process involves a number of different stakeholders. This includes one or more developers who initiate the project, one or more consultants who are responsible for researching potential environmental impacts, and governmental authorities who conduct public hearings, process the reports, and provide a final approval or rejection of the project. The EA process is ideally initiated early when a developer scopes their proposed plans. A final EA report should include the cumulative effects of the project, any public concerns, how to mitigate identified risks, and alternative ways the project could be executed with limited harm to the environment. As such, the type and range of data required for EA reports vary from project to project. One example, described by a stakeholder in our study, is building developments in an inner city area. For such a project, data production can involve recording sounds on-site that can later be used in 'Soundplan', a simulation tool to predict noise levels according to a variation of building structures.

We take inspiration from the principles of case study analysis to perform a more rigorous investigation. A case study analysis "*[...] focuses on understanding the dynamics present within single settings*" [16]. As discussed by Yin [74], case studies are useful for accounting for contextual factors relevant to the phenomenon in which a researcher is interested. Put another way, a case study analysis aims to understand how specific cases operate in their contexts. A case study analysis employs a theoretical lens, here data journeys, to make sense of the examined phenomena [33]. This is useful for our study, as it allows other readers to draw on the theoretical lens we develop to examine other similar phenomena, as well as critically compare this lens to other potential ways of modelling data journey-like activities.

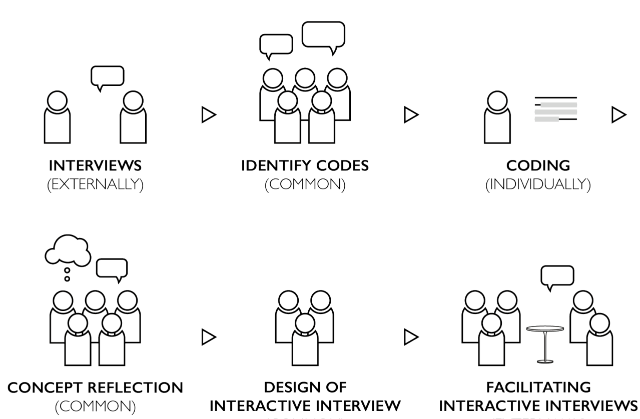

Fig. 1. Overview of the research activities following ten initial interviews.

Table 1. Interview participants and duration of interviews.

Stakeholder type	Organisation	Duration
Consultant 1	NIRAS	46:47
Consultant 2	COWI	31:56
Governmental authority 1	The Danish Environmental Protection Agency	35:66
Governmental authority 2	Danish Energy Agency	47:06
Regional authority 1	Municipality of Esbjerg	59:13
Regional authority 2	Municipality of Aarhus	1:19:18
Developer 1	The Danish Road Directorate	44:40
Developer 2	BaneDK	42:57
NGO 1	The Danish Society for Nature Conservation	16:35
NGO 2	DinGeo	14:25

Case studies have been criticised for sometimes having nontransparent knowledge integration and weak scientific objectivity [58]. To counter this as well as emphasise the value of exploratory case studies, Pauwells and Matthyssens [49] describe a set of principles referred to as the "Four Pillars and the Roof". This consists of 1) theoretical sampling, 2) triangulation, 3) pattern-matching logic, 4) analytical generalisation, and 5) validation through juxtaposition. We utilise these principles by drawing on cases from multiple organisations, relying on the perspectives of different types of stakeholders, conducting the analysis between several researchers, developing the framework iteratively through consultation with previous research and the collected data, and validating the framework through two concrete cases. On the basis of this, the research process followed the structure outlined in Figure 1.

4.2 Interviews with EA Stakeholders

We conducted 10 interviews with different stakeholders involved in conducting and reporting environment assessments. Participants included two engineering consultants, two governmental and two regional authorities, two project developers and two NGO's. An overview of participants and interview duration can be seen in Table 1. The interviews were semi-structured and followed an interview guide with a point of departure in current work practices with a specific focus on navigating environmental data. The guide included questions such as: *"What data do you work on?"*, *"How do you contribute to environmental data?"*, and *"How do you typically search for environmental data?"*. At the end, we asked participants to imagine and describe their practice in 5 and 10 years. Accordingly, the interview guide was a means of encouragement for the participants to give a detailed description of their understanding of environmental data, how data travels, including the challenges meet in this process.

4.3 Analysis and Concept Development

Illustrated in Figure 1, we first identified essential activities of data journeys from Leonelli [34]. The four concepts identified are: Producing, processing, re-purposing, and mobilising. These concepts formed a general theoretical framework on the basis of which we performed a thematic analysis of the interviews. Braun and Clarke describe the deductive orientation in a reflexive thematic analysis as a process where *"theory provides a descriptive lens through which to code and make meaning of the data"* [4]. We used this orientation by performing an initial, individual coding between three researchers. After initial coding, the codes were discussed between these three and two additional researchers through several examples from the data. This solidified the description of each concept

Table 2. Interactive interview participants and duration of interviews.

Stakeholder type	Organisation	Duration
2 consultants	COWI	150 minutes
2 developers	Energinet	180 minutes

in relation to our specific case of EAs. We then sought to ground the concepts further by describing the relations between each concept. We conducted a new analysis of the coded data specifically looking for these relations. An initial round was done by one researcher who analysed all quotes covering two or more codes, followed by a second round of discussion between three researchers.

4.4 Interactive Interviews

To solidify the conceptualisation further, as well as validate the use of the framework within the specific context of EAs, we designed and facilitated two interactive interviews with two different consultancy companies. An overview of participants and duration can be seen in Table 2. In this paper, we focus on the findings from coding the interviews and developing the concepts. We include brief descriptions of these interviews here as further useful insights on the concepts that emerged from discussions with the involved stakeholders.

During the interactive interviews, we asked participants to chronologically describe different data activities throughout three phases of an EA project: 1) Screening, 2) Scoping, and 3) Writing. For each data activity, we asked them to describe involved stakeholders and digital platforms. Then, we asked them to mark how each data activity related to one or more of the four concepts. With this approach, we investigated whether our conceptualisations were useful in the context of EAs, if they supported reflections on the data activities, and if the framework should be expanded, narrowed, or modified in any way. Data collection consisted of note-taking, audio recordings and taking photos during the sessions. We analysed this data by annotating the audio files and discussing the notes between three researchers.

5 FINDINGS

As a part of solidifying the four data activities during our interactive interviews, our participants comfortably recognised the four data activities and instinctively articulated these activities as part of data work in their EA practice. Yet, we also discovered that our participants had a profound desire to view data as static, immutable, and computational bits bringing objectivity and verification to their environmental assessments. This was particularly the case in their attempts to articulate potential solutions to overcome challenges related to how data travel, transform, and create tensions as part of their journey. Instead of embracing the dynamic and fluid materiality of data, they tended to propose solutions in the form of software tools that “tame” the data journey. These imagined software solutions encompassed standardised visions of data management, facilitation of data quality, and smarter digital version control tools. We are not surprised by this due to the high ethical responsibility of EA practitioners where their assessments are expected to rely on objective and sound projections of the impact that human constructions may have on the environment. Given the positive response to the descriptions of the four concepts, we structure our findings according to them, *i.e.*, producing, mobilising, processing, and re-purposing. We summarise the extended data journey framework as an interconnected set of concepts in Figure 2.

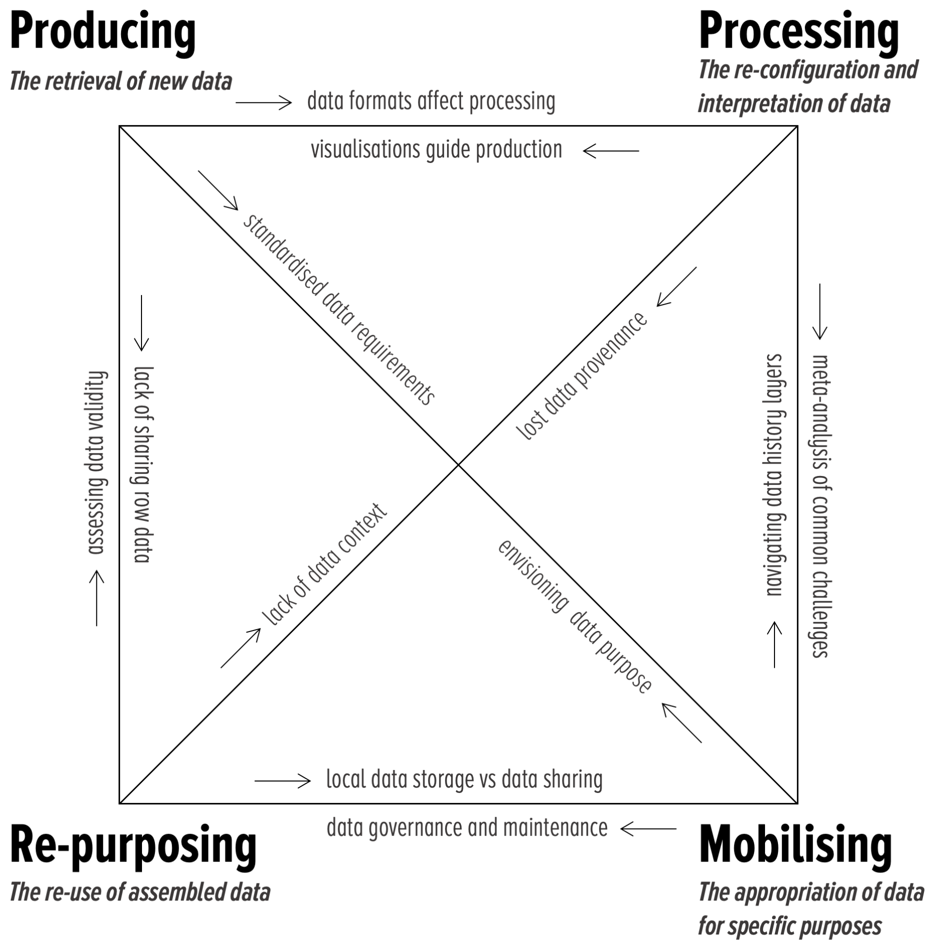

Fig. 2. Data journeys in environmental assessments.

5.1 Producing

We summarise data producing activities as "the retrieval of new data" which on first hand is seemingly simple but, as revealed by our study of EA practice, has many complexities. We show that when data are produced, they are not only brought from the production site to other places but also often transform into new material forms in the process.

For EA practitioners, data producing activities are core to the practice as the final assessments are based on complementary sets of environmental data. Thus, it is crucial that data are represented in reliable formats when produced. One part of data producing activities is locating and deciding data sources and production sites. Data digitisation plays an essential role in these activities; both in terms of digitising analogue data as well as the role of digital platforms used to search for relevant data. For most Danish EA practitioners, producing data starts with a search for and retrieval of environmental data from publicly available, digitised databases that store large quantities of diverse environmental data. This data have already undergone material transformations. One consultant describes how geographic information added to environmental data makes GIS-oriented searches possible, also shaping how they work in these specialised databases: *"Well, we work with some slightly special data. And because of the gravity of what we work with [environmental impact] – at least in a Danish context – we gather data from Miljøportalen* [1]*, ArealInfo* [2]*, and PlanSystem* [3]*. Then there may be some Naturdata* [4]*, etc., but it is most often in GIS data we work with, linked to these different environmental topics"* (Consultant 1). These public databases store various environmental data frequently transformed into standardised, yet specialised, data representations, *e.g.*, GIS (Geographic Information System) data. Thus, determining what data are relevant to retrieve is partly shaped by the kinds of available data, data formats, data models and data representations; *"... we used to have a GIS solution when working with large facilities, i.e., not on a 3D but on a 2D model, where you insert the alignment of a road or a railway. You could fetch all sorts of things imaginable from the 'background data' from this."* (Developer 2).

Although large quantities of data can be retrieved through public databases, there are cases where new data have to be produced. In EA practice, this commonly means to produce new environmental data from specific sites affected by a potential development project, *e.g.*, measuring sound or counting endangered species in location. In those data-producing activities, EA practitioners must ensure that produced data adequately represent the data source at the production site *e.g.*, bio-diversity or human health concerns. Tensions may emerge in those activities, as the amount and what data sources should be included or excluded are situated to cost-effective measures related to specific EA cases; "*here, there are some big, old trees that bats may not inhabit, but there could be bats. So how do we handle it through an assessment without having to go out and make a lot of data?*" (Regional authority 2). To help eliminate possible "human-triggered" data frictions as data travels from production sites to specialised platforms, EA practitioners may make use of digital tools. One developer describes "*...it will not be long until we change our GIS platform. It will provide some new tools also in terms of data collection when we are out in the field which will be easier. So I think today it's a bit of a hassle that you collect in a way and you have to have data converted over to a database. There are too many hands to cross. A lot of work is being done to make it more smooth*" (Developer 1).

These digital tools may also play a crucial role in how EA practitioners validate their produced data. Some tools embed standardised data algorithms that materialise governmental guidelines and regulations shaping both what data are produced and expectations of data validity in EA practice: *"... there are also standardised tools for noise calculation, which are always used. And then, there are regulations and instructions from the Danish Environmental Protection Agency, such as the limits for exposure values of noise. I also see these instructions as tools - an assessment criteria for how the noise types are settled upon"* (Developer 2).

The above findings illustrate that digital data, in practice, are not static, isolated entities but rather highly dynamic elements that, when produced, also integrate into different aspects of data work practices. In this, the digital tools and methods from which data are produced evolve over time, influencing the social-material conditions of working in data-intensive practices. Furthermore, as data producing activities transform data, they create tensions as data sources are to be included or excluded and need validating based on competencies and expectations embedded in practice.

[1] https://miljoeportal.dk: A public website with various databases of environmental data, infrastructures and services aiming to support digital environmental management.
[2] https://arealinformation.miljoeportal.dk/html5/index.html?viewer=distribution: A specialised GIS database tool to search for spatial GIS data.
[3] https://planinfo.erhvervsstyrelsen.dk/plandatadk: A digital register for spatial planning in Denmark. The tool ensures that spatial planning data is unique and digitally accessible.
[4] https://naturdata.miljoeportal.dk/: A digital search register for species found in connection with Danish authorities' nature monitoring.

5.2 Mobilising

We summarise data mobilising activities as "the appropriation of data for specific purposes". To mobilise data, practitioners transform them, merge them with other data, and change their materiality. Thus, data mobilisation means that data travel as they undergo modifications to conform to new sites and circumstances. Our analysis illustrates how this plays out in EA practices.

As data are core materials in forming environmental assessments, mobilising data for specific projects is essential in EA practice. One consultant explains that "*... working in the planning system with these data models, you decide for yourself what kind of program you want to use–and you also decide for yourself how you to put it together*" (Consultant 1). Specialised tools are frequently used when appropriating data for particular purposes. Here, EA practitioners often experiment and learn through sense-making experiences shaped by the digital and its material representation. One consultant states that "*...in our practice, we use the 3D Studio MAX program, where you transform this CAD material to some visualisations. Those visualisations are often geographically measured [data] - and they are super accurate in some geographical grit–so you can say that if you had an image, it would be a piece of cake to link it geographically to a place in reality*" (Consultant 1). Furthermore, mobilising data in those activities involves appropriating data representation as sense-making mechanisms to understand, explain, and make data available internally in a team and to external partners. The other interviewed consultant stated that "*Sometimes, it's nice just to have something on a piece of paper or a blackboard, which might be too fancy to make into 3D. But it could be fun to look at what you can do with 3D in an environmental assessment and explain the project to those who have to do an environmental assessment and the authorities*" (Consultant 2).

In data mobilising activities, data often arrive in the hands of a practitioner in complex, digitalised data formats (*e.g.*, CAD models), requiring domain-specific competencies to then appropriate data using these tools (*e.g.*, CAD Tools). One consultant describes that "*... SoundPLAN is used for different things, e.g., company noise, road noise. There are also wind turbines which have their own logic, where you work in something called WindPRO which also delivers noise cards, but which looks a little different*" (Consultant 1). As data mobilise, they travel and transform through digital tools that also continuously change over time. This mobilisation challenges the provenance of data (*e.g.*, data history and place of origin) and hence their continued usability as described by the local authority: "*...and there is another problem, typically when you set up a model in connection with noise or groundwater. The software that is used continuously changes – so after a few years, you will not be able to use the data*" (Regional authority 1). Moreover, as data continuously are mobilised, add complexity and transparency issues to their transformed materiality, making data provenance difficult to verify in practice. The governmental authority explains that "*...[the GIS system] often has several layers [of data]–and you're not necessarily in the wrong layer. But with more layers, it can sometimes be a bit messy figuring out what layer is the newest and most current [data set]*" (Governmental authority 1). When data mobilisation activities transform data and transparency is lost, tensions regarding how the argumentation in EA assessments came about can surface when regarded by other stakeholders. An NGO explains that "*...we simply complain if we believe that they [environmental assessment reports] have not been prepared well enough and if we believe that there is a lack of a basis, or if they are not exhaustive enough to make an assessment*" (NGO 1).

The above findings illustrate that data mobilising activities in EA practice transform data, as EA practitioners use different digital tools in their appropriation of data to be able to experiment and reason for their assessment. Moreover, data mobilisation can create tensions related to data provenance as transparency becomes more complex to maintain and practitioners attempt to make sense of transformed data layers.

5.3 Processing

We summarise data processing activities as "the re-configuration and interpretation of data". Data processing activities illustrate that data are dynamic elements capable of creating new socio-material configurations as they journey, such as their ability to aggregate, compute and be re-interpreted in different formats (*e.g.*, words and images). Our analysis illustrates how this plays out in EA practices.

One part of processing environmental data in EA activities is to create meaningful data configurations allowing for different data interpretations. Data visualisations and practitioners' ability to compute, aggregate, and visualise patterns are pivotal here. One of the interviewed developers explains that "*...SoundPLAN is a computer tool for acoustic models where you can see how noise from a noise source spreads in the landscape. Here you simply enter traffic noise, a 3D model of the landscape and the facility you want to build, – and then you can simply [visually] evaluate how the noise spreads in the landscape.*" (Developer 2). Similarly, EA practitioners use increasingly complex environmental data configurations to help predict future environmental impact scenarios. As the other interviewed developer describes: "*...so, we evaluate how the noise will be in the future, by computing it. And the tools have become pretty good for this. And because computing power has improved so much, you can put in a 3D model of a railway of 20 km and see how it continues.*" (Developer 2). When data are processed, specialised and digital tools become significant elements of data work practices calling for specialised competencies to make data re-configurations. At the same time, EA practitioners also use these tools to shape a shared social consensus around the interpretive data work in these practices. This was described by one consultant: "*...you often receive anything from a technical specification from a supplier to some technical element of a water purification system – where you then have to transform it into something that is more understandable.*" (Consultant 1).

Data processing activities can produce advanced data visualisations of predicted future scenarios that may facilitate more embodied data practices. These embodied data experiences allow EA practitioners to visualise and form new interpretations of their assessment. One consultant states that "*...you may take pictures in 360 degrees, which you then can put your wind farm into it. You can actually sit down, and use 3D glasses to stand in this place. And if it is down by the sea, i.e. by the coast – then you can go down there and put on the 3D glasses, and see what this wind farm will look like here*" (Consultant 2). As environmental data are processed, they typically materialise as re-interpreted words and visual images in an EA report. However, sometimes more interactive visualisations are used to make processed data more meaningful to a broader audience. A developer describes this: "*...instead of having them [noise maps of noise calculations] in the report, they are put on a digital map on a website where it is possible to zoom in and out if something is of interest to you. For example, if you are a citizen, then you might be interested in knowing how the noise will be where you live.*" (Developer 2). Furthermore, processing data is often done on a meta-level (*e.g.*, sound measures re-configured into visualisations) that may lead to data formats which rely on certain standard data-processing methods and tools (noise calculations in proprietary software tools, e.g., SoundPLAN), needing specialised competencies (being able to work in, e.g., SoundPLAN). In these activities, tensions may emerge concerning how to obtain a collective consensus regarding how practitioners report on processing-data activities. A consultant states that "*I believe in making an environmental assessment crystal clear; you should manage to use the proper argumentation and be able to show it. Just like in a scientific article, where you have some proper references and explain it, you could say. And here, data are – not just our own but also people's data – important.*" (Consultant 1).

Our findings on data processing activities in EA practice demonstrate that transforming digital data into more meaningful configurations shapes the ways they are interpreted. Accordingly, our findings show that the rich material characteristics of digital data make them prone to a large variety of interpretations. In practice, this data transformation is co-performed by EA practitioners and their specialised software tools and infrastructures. Yet, tensions may emerge on how practitioners maintain and evolve the qualification of intuitive, collective understandings of processing data.

5.4 Re-purposing

We summarise data re-purposing activities as "the re-use of assembled data". Despite data often coming in complex forms and formats, data still journey in these re-purposing activities. Moreover, since digital data are usually characterised by their longevity, well-managed and shareable data have the potential to be re-purposed. Our analysis illustrates how this plays out in EA practices.

While EA practitioners base their assessments primarily on a diverse set of environmental data, they also complement their assessment with existing data, interpretations and knowledge shared by other EA practitioners. This way of re-purposing data is very similar to how scientific practitioners present their work in scientific articles where overall arguments are re-used but the raw data sources are typically not disclosed. One consultant states that "*...you also use these environmental reports a bit like a scientific article as references. Yes, it is very similar to the structure of an article where I can read about the data set. But I can not really use the set in my own calculations if I cannot get it from those who have made it*" (Consultant 1). When data are re-used in re-purposing activities, assembled data are commonly acknowledged in the same way scientific practitioners re-purpose scientific knowledge via a citation to a journal article. Hence, the output of one environmental assessment project may very well be an input of another. This way of accumulating data and knowledge can challenge practitioners because data contexts can be difficult to retain. Additionally, there is an inherent conflict in that data are typically stored locally by consultants for easy internal access and not by the public authorities who instituted the collection. Such tensions raises questions concerning the governance of data, calling for principles outlining the organisation of data access. The interviewee from the regional authority describes that "*In environmental assessments investigations, you produce a lot of data. Those data are only stored with the consultants, cannot be shared and are lost. It's crazy that our organisation pay vast amounts of money to a consultant who produces something and writes an EA, and then other authorities cannot use the same data for subsequent work.*" (Regional Authority).

Similarly, assembled EA data formats, typically EA reports, are often materialised in rigid formats (*e.g.* pdf, word docs, images, and even printed formats), making it difficult or impossible to conduct advanced searches. Furthermore, there is currently no centralised Danish repository to store and search for previous reports making it challenging for practitioners to locate existing data and knowledge. One consultant describes that "*...I often try to find something similar to what I'm working with because then I can assess if others have found something [that can have an impact]. And I'm often challenged by the fact that they [other reports] just lie there purely analogue [pdf format]–if you can find the report at all.*" (Consultant 1). Yet, for some practitioners, managing and revising environmental data in shared national data repositories are part of their practices, as is described by a consultant: "*...some have agreements with Miljøportalen that you put data back when you do field studies. Not all developers will pay for this, as there is an economic element in providing something and getting something in return. It is most often at Vejdirektoratet and BaneDK, i.e., larger projects were reporting back to the system, that is part of it. But the thing that is being reported is very much incomplete and unorganised*" (Consultant 1).

Our findings on re-purposing activities show these data are re-used frequently as data work from prior projects is used to accumulate knowledge and bring quality to EA practice. For some practitioners, sharing well-managed data are part of their practice, while others share data and re-purposed data evaluations for transparency reasons. Nonetheless, retaining data context is challenging as provenance is often lost when data are assembled, and raw data sources are inaccessible. Tensions call for governing principles to ensure re-purposed data can be shared among practitioners.

6 DISCUSSION

Our findings show that applying the four data journey activities as an analytical lens to understand how data journey can lead to rich and detailed descriptions of data work practice. Furthermore, the findings enable us to point to tensions that can be addressed to support future practice. We summarise data journey activities and emergent tensions in Figure 3. We identify two main implications for further research based on these tensions. First, they illuminate opportunities and challenges for EA practice. Each tension can be unfolded in future research with a focus on how to cultivate, reflect on, or potentially resolve them. Second, the tensions reflect that by using data journeys as an analytical lens, we are able to identify and describe conflicting perspectives and frictions in environmental assessment practices. We speculate a similar outcome for other domains because of a shift in focus from isolated interactions between a practitioner and data, towards framing data journey as entanglements and frictions between data and data work activities as they travel.

In the following subsections, we synthesise our findings in relation to previous research and present three main aspects in which this paper adds to that research by 1) describing the merits of using data journey theory as a unit of analysis through a concrete case, 2) outlining resulting data tensions arising from that analysis, and 3) discussing how the developed framework can be used to create a foundation for future design work to support data journey activities. Each section is summarised in a list of main takeaways. We acknowledge that these contributions emerge from a study that is anchored in the domain of environmental assessment practice. We argue, however, that the generalised nature of the findings points to its usefulness across other data work practices.

6.1 Data Journey Activities as a Unit of Analysis

Our findings show that digital data in EA practice are dynamic elements that shape practice and, in turn, become fluid objects in constant transformation. For example, the two consultants in our study described the required competencies to utilise a range of tools for processing and mobilising data. As such, our findings illustrate that EA practitioners often rely on continuous interpretations and re-configurations of data to form environmental assessments that in turn shape EA practice. Thus, rather than seeing data as static, immutable, objective, and detached, computational "raw" bits as input to data-intensive work practices [19], our aim has been to

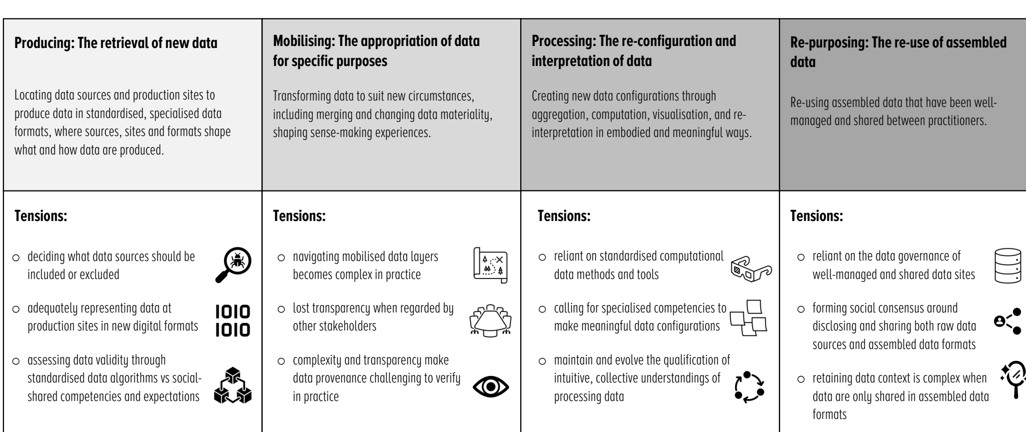

Fig. 3. Data journey activities and emergent tensions in environmental assessments.

capture entanglements of data practices by making data and their interactive [22, 65, 72, 73] and socio-material elements [10, 12, 45, 62], our unit of analysis.

Other research has conceptualised the ways in which different data practitioners collaborate in data work [5, 7, 11, 39, 60, 75], data workers' interventions [41, 43], and particular aspects of data frictions [42, 47, 56]. While these discussions can be fruitful for understanding collaborative and diverse aspects of data work [5], we believe that our conceptualisation and detailing of data journey activities bring alternative perspectives to understanding data work as a socio-material practice [45], which also captures the varieties of different contexts and particularities of work domains. As our analysis of data activities in EA practice demonstrates, we are able to capture the dynamic entanglements and tensions of data transformations and data work practices, by following how data journey through space (physical and virtual), time, and social situations. We believe this opens new and alternative ways to understand the "*open issues in the behavioural and social study of data science and data science workers*" [40].

Building on previous research exploring data mapping [1, 61, 68], we argue that analysing data journeys through the four data activities brings a complementary perspective to understanding the dynamic entanglements between the digitisation of data [10, 55] and data work practices [44, 62]. We extend and present a conceptual framework of four data-intensive activities, *i.e.*, producing, mobilising, processing, and re-purposing. While our conceptualisation of the four data journey activities is inspired by Leonelli's [34] abstract, yet cohesive data journey definition, we do not intend to claim that these four data activities are exhaustive to the routines of EA practitioners. In fact, our findings also show that these activities overlap, *e.g.*, when processing is carried out as a means to mobilise data. Nevertheless, we believe that our articulation and more in-depth definitions of the four data activities solidify and complement perspectives on the prior data journey framings of both Leonelli [34] and Bates et al. [1], and other descriptions of data activities [5, 23]. While the participants in our study work intensively with environmental data to create objective, verifiable, transparent, and comprehensive environmental assessments, our findings suggest that social elements, (*e.g.*, competencies, meanings, and other materials [62]) also shape data, data activities, and not the least, data tensions. This was, *e.g.*, the case when assessing whether data can be re-purposed, as detailed by the regional authority who had to communicate with developers to access their data. Furthermore, our findings show how other materials (*e.g.*, digital tools, infrastructures, and platforms) embedded in EA work practices shape the ways that data journey. Data work in EA practices is guided by the practitioners' domain-specific competencies [36], while digital tools and infrastructures operate as "co-performers" [28] to support practitioners in bringing about more meaningful data interpretations [32, 36, 38, 43]. This, *e.g.*, includes how EA practitioners process environmental data to be re-interpreted via interactive technology to support a more meaningful embodied experience of how a construction may impact the environment. By consolidating the four data journey activities into a coherent analytical framework and using that in the context of the data work practice of environmental assessment, we illustrate how this framework gives a nuanced lens for understanding data dynamics and how they shape practice.

Key takeaways:

- ***Framing*** *data journey activities as a unit of analysis reorients our focus from individual practitioner experiences with data to understanding the dynamics of data transformation resulting from how practitioners engage and disengage with data in their work practices.*
- ***Solidifying*** *the four data journey activities, enable us to describe how different data journey activities relate and are entangled in environmental assessment practice. We argue that this is transferable to investigations of other data-intensive practices.*

6.2 Accounting for Data Tensions

Similarly to Desjardin et al.'s [8] study of IoT data in domestic environments, our findings reveal the complexity of capturing entanglements of the objective notions and subjective understandings of data work. Using a lens on those data entanglements in our analytic work, we also uncovered tensions in these data work practices. Similarly to Pink et al. [52], Muller and Strohmayer [42], and Edwards et al. [15], our findings show that tensions, conflicts, and frictions emerge at a data level as part of interwoven data journey activities. For example, in order to effectively re-purpose environmental data in EA practice, data provenance needs to be maintained. Nonetheless, data provenance is partly influenced by mobilising data activities, either through adding new material layers to the data or performing new computations on existing data to meet current standards. Another example is the maintenance of data context, which is essential in mobilising data to ensure they can be re-purposed in future projects. We found a challenge for both tensions related to the collective problem of data governance [2] where action principles may guide *e.g.*, who is accountable for maintaining data assets. Our finding shows that when EA data are locally stored, such maintenance is simple, but when the EA data are collectively shared, it becomes more complex to maintain records that trace all data transformations and contexts.

Comparing our findings to prior research [5, 15, 34, 41, 43, 46, 75], EA data practices resemble other data-intensive practices. However, by making data journeys our unit of analysis and framing data work as a socio-material practice [45], we are able to show that tensions emerge as part of data work practice. For instance, our findings reveal that EA practitioners, by processing data, will aggregate, compute, visualise and re-interpret data that also shape their accounting of the final assessments. Nevertheless, it is also paramount for these practitioners to preserve data transparency when assessing data validity, as inadequate environmental assessments can have grave environmental consequences. Furthermore, when those tensions are articulated and acted upon in EA practice, we demonstrate that data transform into other material forms, re-configuring different meanings, and possibly needing additional competencies to manage them in practice. Hence, we argue that by accounting and articulating tensions by analysing data journey activities, we are able to illustrate the variety of meanings and dynamics of data's materiality as tensions are reflected and acted upon (or not) in practice [15]. Being able to articulate and account for tensions as an emergent aspect of data work we extend prior data journey framings from Bates et al. [1] and Leonelli [34].

We summarised the identified emergent tensions in EA practice in Figure 3. We do not intend to claim that these tensions are exhaustive or conclusive, as others show how tensions in data work, *e.g.*, data transparency [18], data trust [47], data provenance/history [27], data negotiations [60], data governance [2], and data sensemaking [43] play key roles in narrating the dynamic and situational essence of data [5, 17]. For instance, re-purposing is of high value in EA practice, both when the data are locally stored and when produced by other consultants, authorities, or public platforms. However, while consultants describe a strong social consensus of wanting to share data with others, they are also in direct competition. This may create tensions because EA practitioners depend on each other to effectively gather and validate data necessary for the assessment. Not having access to shared data knowledge influences both re-purposing and producing activities, as well-managed and shared data sites shape those activities. Another example is that for producing data, as this EA activity typically involves methods that are standardised after governmental requirements. Sometimes these standards are embedded as algorithms in digitalised tools, and sometimes specialised competencies to the disciplines are involved. These methods result in data formats requiring standardised computational data methods and tools to process, then again calling for specialised competencies to make meaningful data configurations. By uncovering these tensions through the four data journey activities, alternative perspectives on the dynamics of tensions emerge, which may enable practitioners outside the research field to articulate and account for how tensions are present and potentially acted upon in their daily work practices.

Key takeaways:

- ***Tensions*** *are a useful way to describe socio-material entanglements that are uncovered through the data journey lens.*
- ***The dynamics*** *of data materiality and data work is uncovered when analysing how practitioners act on data tensions.*

6.3 Creating a Foundation for Designing Data Journeys

The research presented in this paper is a step towards expanding understanding and detailing data journeys in particular domains requiring expert knowledge. While previous work proposes that data creation [17], data stories [68], data mappings [1], data rhetoric [27], more-than-human data stories [35], and data (service) design [61] can be used to foreground data in design activities, our work shows that a data journey lens helps to capture the dynamics and entanglements of data, their materiality, and the practices in which they are embedded. As other studies illustrate (*e.g.*, [61, 68]), it can be challenging for practitioners to articulate data as fluid, pluralistic elements of practice and bring that perspective into the imagination of future designs.

We believe that foregrounding data journeys via our framework and putting a focus on data transformation as part of data-intensive practices can be used as a point of departure for future design activities. As our findings show, other aspects of practices, such as competencies, meanings and other materials, are also articulated and detailed as part of this analytic work. The EA practitioners in our study used digital tools and data (materials) as well as specialised competencies to experiment and make meaning of data in the situation as part of their EA practice. Argued by Leonelli [34], foregrounding data illustrates that data practices and transforming data into knowledge is not a matter of finding "*generalist algorithms, clustering methods, robust infrastructure and/or clever apps*" but rather points to directions that better encounter practitioners' adaptive, creative, and ever-changing data practices when aiming to design data services, infrastructures and digital tools that aim to support data workers [34]. Thus, we suggest that recognising the dynamics of data journeys through design activities may better support practitioners' creative experimentation and embodied sense-making of data in practice. One suggested step in this direction could be to use data and "*the roles of non-human material elements or what we refer to as things in everyday domestic practice*" [70] as a point of departure for design that broader embraces materials, competences, and meanings of practice as a unit for design [25, 29].

A way forward could be to view digitalisation as a matter of designing where; 1) digital tools are viewed as work companions or non-human "co-performers" [28] of practice and 2) data are recognised as fluid, material "things" [24, 66, 67, 70] that continuously shape work practices. This could enrich and shape meaningful data practices rather than viewing the design of innovative digital technologies as a goal in itself [55]. While this could be carried out with a variety of approaches, we suggest that data journey activities can play a central role in such design processes, *e.g.*, for acting out tensions in data work to cultivate reflections about them or for articulating how data-intensive practices change with continuous digitisation. We believe data journeys can help bring such alternative ways of thinking about design as part of data-intensive practices.

Key takeaways:

- ***Activities*** *of data journeys are a valuable point of departure to support practitioners in articulating the roles data play in their work practices.*

- *A **data journey** lens identifies data, tools, infrastructures and other materials, meanings, and competencies used to engage and disengage with data in practice, the knowledge of which can inform subsequent design activities.*

7 CONCLUSION

In this paper, we report on an investigation into using data journey activities as a unit of analysis in data-intensive practice. We develop an extension of existing theory in the form of a theoretical framework through an analysis of interviews with practitioners in the domain of environmental assessment. With a point of departure in Leonelli's [34] theory of data journeys, we show that this lens can be further extended to shed light on tensions for further design of support of data journey activities. Through the articulation and development of four data activities – producing, mobilising, processing, and re-purposing – we illustrate that by making data, including their socio-materiality and how they journey in practice, our unit of analysis, we are able to capture social, temporal, and spatial entanglements of data practices. We present two main contributions: 1) The theoretical framework to be used as a unit of analysis, and 2) a set of tensions anchored in EA practice that form a foundation for future design activities in this area. While this study was anchored in EA practice, we speculate that the theoretical framework can be extended to other domains and used to enable practitioners to articulate the complexities of data journey activities in their work.

ACKNOWLEDGMENTS

This project has received funding from Innovation Fund Denmark - Grant agreement number, 0177-00021B DREAMS. We thank the participants for their time and engagement.

REFERENCES

[1] Jo Bates, Yu-Wei Lin, and Paula Goodale. 2016. Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society* 3, 2 (2016), 2053951716654502. https://doi.org/10.1177/2053951716654502

[2] Olivia Benfeldt, John Stouby Persson, and Sabine Madsen. 2020. Data governance as a collective action problem. *Information Systems Frontiers* 22, 2 (2020), 299–313. https://doi.org/10.1007/s10796-019-09923-z

[3] Susanne Bødker. 2006. When Second Wave HCI Meets Third Wave Challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (Oslo, Norway) *(NordiCHI '06)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/1182475.1182476

[4] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis - A Practical Guide*. Sage Publications Ltd.

[5] Samuelle Carlson and Ben Anderson. 2007. What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication* 12, 2 (01 2007), 635–651. https://doi.org/10.1111/j.1083-6101.2007.00342.x arXiv:https://academic.oup.com/jcmc/article-pdf/12/2/635/22317230/jjcmcom0635.pdf

[6] Sumandro Chattapadhyay. 2014. Access and Use of Government Data by Research and Advocacy Organisations in India: A Survey of (Potential) Open Data Ecosystem. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV '14)*. Association for Computing Machinery, New York, NY, USA, 361–364. https://doi.org/10.1145/2691195.2691262

[7] Joshua Ddamba and Yvonne Dittrich. 2015. Data for Participation and Participation as Data: Supporting Incremental Participatory Decision-Making in Urban Planning. *Selected Papers of the IRIS, Issue Nr 6 (2015)* (2015).

[8] Audrey Desjardins, Heidi R. Biggs, Cayla Key, and Jeremy E. Viny. 2020. *IoT Data in the Home: Observing Entanglements and Drawing New Encounters*. ACM, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376342

[9] Grace K.C. Ding. 2008. Sustainable construction—The role of environmental assessment tools. *Journal of Environmental Management* 86, 3 (2008), 451–464. https://doi.org/10.1016/j.jenvman.2006.12.025

[10] Paul Dourish. 2017. *The stuff of bits: An essay on the materialities of information*. MIT Press.

[11] Paul Dourish, John Lamping, and Tom Rodden. 1999. Building Bridges: Customisation and Mutual Intelligibility in Shared Category Management. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work* (Phoenix, Arizona, USA) *(GROUP '99)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/320297.320299

[12] Paul Dourish and Melissa Mazmanian. 2013. Media as Material: Information Representations as Material Foundations for Organizational Practice. *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies* 3 (2013), 92.

[13] Christof Ebert and Carlos Henrique C. Duarte. 2018. Digital Transformation. *IEEE Software* 35, 4 (2018), 16–21. https://doi.org/10.1109/MS.2018.2801537

[14] Paul N Edwards. 2010. *A vast machine: Computer models, climate data, and the politics of global warming.* Mit press.

[15] Paul N Edwards, Steven J Jackson, Geoffrey C Bowker, and Cory Philip Knobel. 2007. Understanding infrastructure: Dynamics, tensions, and design. (2007).

[16] Kathleen M Eisenhardt. 1989. Building theories from case study research. *Academy of management review* 14, 4 (1989).

[17] Melanie Feinberg. 2017. A Design Perspective on Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2952–2963. https://doi.org/10.1145/3025453.3025837

[18] Asbjørn William Ammitzbøll Flügge, Thomas Hildebrandt, and Naja Holten Møller. 2020. Algorithmic Decision Making in Public Services: A CSCW-Perspective. In *Companion of the 2020 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '20)*. Association for Computing Machinery, New York, NY, USA, 111–114. https://doi.org/10.1145/3323994.3369886

[19] Lisa Gitelman. 2013. *Raw data is an oxymoron.* MIT Press, Cambridge, Mass.

[20] Karey Helms. 2019. Do You Have to Pee? A Design Space for Intimate and Somatic Data. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1209–1222. https://doi.org/10.1145/3322276.3322290

[21] Hiroshi Ishii, Dávid Lakatos, Leonardo Bonanni, and Jean-Baptiste Labrune. 2012. Radical Atoms: Beyond Tangible Bits, toward Transformable Materials. *Interactions* 19, 1 (Jan. 2012), 38–51. https://doi.org/10.1145/2065327.2065337

[22] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, Atlanta Georgia USA, 234–241. https://doi.org/10.1145/258549.258715

[23] Azra Ismail and Neha Kumar. 2018. Engaging Solidarity in Data Collection Practices for Community Health. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 76 (nov 2018), 24 pages. https://doi.org/10.1145/3274345

[24] Rikke Hagensby Jensen, Enrique Encinas, and Dimitrios Raptis. 2022. Spicing It Up: From Ubiquitous Devices to Tangible Things Through Provocation. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Daejeon, Republic of Korea) *(TEI '22)*. Association for Computing Machinery, New York, NY, USA, Article 33, 15 pages. https://doi.org/10.1145/3490149.3502257

[25] Rikke Hagensby Jensen, Yolande Strengers, Dimitrios Raptis, Larissa Nicholls, Jesper Kjeldskov, and Mikael B. Skov. 2018. Exploring Hygge As a Desirable Design Vision for the Sustainable Smart Home. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. ACM, New York, NY, USA, 355–360. https://doi.org/10.1145/3196709.3196804

[26] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. 2011. Wrangler: Interactive Visual Specification of Data Transformation Scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3363–3372. https://doi.org/10.1145/1978942.1979444

[27] Vera Khovanskaya and Phoebe Sengers. 2019. Data Rhetoric and Uneasy Alliances: Data Advocacy in US Labor History. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1391–1403. https://doi.org/10.1145/3322276.3323691

[28] Lenneke Kuijer and Elisa Giaccardi. 2018. Co-Performance: Conceptualizing the Role of Artificial Agency in the Design of Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 125, 13 pages. https://doi.org/10.1145/3173574.3173699

[29] Lenneke Kuijer, Annelise de Jong, and Daan van Eijk. 2013. Practices as a Unit of Design: An Exploration of Theoretical Guidelines in a Study on Bathing. *ACM Trans. Comput.-Hum. Interact.* 20, 4, Article 21 (Sept. 2013), 22 pages. https://doi.org/10.1145/2493382

[30] Peter Kun, Ingrid Mulder, Amalia de Götzen, and Gerd Kortuem. 2019. Creative Data Work in the Design Process. In *Proceedings of the 2019 on Creativity and Cognition* (San Diego, CA, USA) *(C&C '19)*. Association for Computing Machinery, New York, NY, USA, 346–358. https://doi.org/10.1145/3325480.3325500

[31] Kari Kuutti and Liam J. Bannon. 2014. The turn to practice in HCI: towards a research agenda. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM Press, New York, New York, USA, 3543–3552. https://doi.org/10.1145/2556288.2557111

[32] Chin Loong Law, Paul Roe, and Jinglan Zhang. 2012. Using Mobile Technology and Augmented Reality to Increase Data Reliability for Environmental Assessment. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia) *(OzCHI '12)*. Association for Computing Machinery, New York, NY, USA, 327–330. https://doi.org/10.1145/2414536.2414589

[33] J. Lazar, J. H. Feng, and H. Hochheiser. 2017. *Research methods in human-computer interaction.* Morgan Kaufmann.

[34] Sabina Leonelli. 2020. Learning from data journeys. In *Data journeys in the sciences*. Springer, Cham, 1–24.
[35] Deborah Lupton, , and Ash Watson. 2021. Towards more-than-human digital data studies: developing research-creation methods. *Qualitative Research* 21, 4 (2021), 463–480. https://doi.org/10.1177/1468794120939235
[36] Ivar Lyhne and Lone Kørnøv. 2013. How do we make sense of significance? Indications and reflections on an experiment. *Impact Assessment and Project Appraisal* 31, 3 (2013), 180–189. https://doi.org/10.1080/14615517.2013.795694
[37] N. Mahyar, D.V. Nguyen, M. Chan, J. Zheng, and S.P. Dow. 2019. The Civic Data Deluge: Understanding the Challenges of Analyzing Large-Scale Community Input. In *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*. 1171–1181. https://doi.org/10.1145/3322276.3322354
[38] Rohit Mehra, Vibhu Saujanya Sharma, Vikrant Kaulgud, and Sanjay Podder. 2019. AssessAR: An Augmented Reality Based Environmental Impact Assessment Framework. In *25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 104, 2 pages. https://doi.org/10.1145/3359996.3365034
[39] Naja Holten Møller, Claus Bossen, Kathleen H. Pine, Trine Rask Nielsen, and Gina Neff. 2020. Who Does the Work of Data? *Interactions* 27, 3 (apr 2020), 52–55. https://doi.org/10.1145/3386389
[40] Michael Muller, Melanie Feinberg, Timothy George, Steven J. Jackson, Bonnie E. John, Mary Beth Kery, and Samir Passi. 2019. Human-Centered Study of Data Science Work Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299018
[41] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300356
[42] Michael Muller and Angelika Strohmayer. 2022. Forgetting Practices in the Data Sciences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 323, 19 pages. https://doi.org/10.1145/3491102.3517644
[43] Gina Neff, Anissa Tanweer, Brittany Fiore-Gartland, and Laura Osburn. 2017. Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science. *Big Data.* 5, 2, Article 21 (2017), 85-97 pages. https://doi.org/10.1089/big.2016.0050
[44] Wanda J. Orlikowski. 2000. Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science* 11, 4 (2000), 404–428. https://doi.org/10.1287/orsc.11.4.404.14600
[45] Wanda J. Orlikowski. 2007. Sociomaterial Practices: Exploring Technology at Work. *Organization Studies* 28, 9 (2007), 1435–1448. https://doi.org/10.1177/0170840607081138
[46] Samir Passi and Steven Jackson. 2017. Data Vision: Learning to See Through Algorithmic Abstraction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2436–2447. https://doi.org/10.1145/2998181.2998331
[47] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 136 (nov 2018), 28 pages. https://doi.org/10.1145/3274405
[48] Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society* 7, 2 (2020), 1–13. https://doi.org/10.1177/2053951720939605
[49] Pieter Pauwels and Paul Matthyssens. 2004. The Architecture of Multiple Case Study Research in International Business. *Handbook of Qualitative Research Methods For International Business* (2004), 125–143.
[50] James Pierce, Yolande Strengers, Phoebe Sengers, and Susanne Bødker. 2013. Introduction to the Special Issue on Practice-Oriented Approaches to Sustainable HCI. *ACM Trans. Comput.-Hum. Interact.* 20, 4, Article 20 (Sept. 2013), 8 pages. https://doi.org/10.1145/2494260
[51] Kathleen H. Pine and Max Liboiron. 2015. The Politics of Measurement and Action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3147–3156. https://doi.org/10.1145/2702123.2702298
[52] Sarah Pink, Minna Ruckenstein, Robert Willim, and Melisa Duque. 2018. Broken data: Conceptualising data in an emerging world. *Big Data & Society* 5, 1 (2018), 2053951717753228. https://doi.org/10.1177/2053951717753228
[53] Carlos Gerardo Prieto-Alvarez, Roberto Martinez-Maldonado, and Simon Buckingham Shum. 2018. Mapping Learner-Data Journeys: Evolution of a Visual Co-Design Tool. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) *(OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 205–214. https://doi.org/10.1145/3292147.3292168
[54] Andreas Reckwitz. 2002. Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory* 5, 2 (2002), 243–263. https://doi.org/10.1177/13684310222225432

[55] João Reis, Marlene Amorim, Nuno Melão, and Patrícia Matos. 2018. Digital Transformation: A Literature Review and Guidelines for Future Research. In *Trends and Advances in Information Systems and Technologies (WorldCIST'18)*, Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis, and Sandra Costanzo (Eds.). Springer International Publishing, Cham, 411–421. https://doi.org/10.1007/978-3-319-77703-0_41

[56] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone Wants to Do the Model Work, Not the Data Work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. https://doi.org/10.1145/3411764.3445518

[57] Pedro Sanches and Barry Brown. 2018. Data Bites Man: The Production of Malaria by Technology. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 153 (nov 2018), 19 pages. https://doi.org/10.1145/3274422

[58] Marja Sarvimaki. 2017. Case study strategies for designers: Teaching integrative data evaluation. In *EAAE/ARCC 2016 International Architectural Research Conference: Architectural research addressing societal challenges*. CRC Press.

[59] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2018. Data work in a knowledge-broker organisation: how cross-organisational data maintenance shapes human data interactions. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. 1–12. https://doi.org/10.14236/ewic/HCI2018.14

[60] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Co-Designing Data Experiments: Domain Experts' Exploration and Experimentation with Self-Selected Data Sources. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 20, 11 pages. https://doi.org/10.1145/3419249.3420152

[61] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Foregrounding data in co-design – An exploration of how data may become an object of design. *International Journal of Human-Computer Studies* 143 (2020), 102505. https://doi.org/10.1016/j.ijhcs.2020.102505

[62] Elizabeth Shove, Mika Pantzar, and Matt Watson. 2012. *The dynamics of social practice: Everyday life and how it changes*. Sage.

[63] Erik Stolterman and Anna Croon Fors. 2004. *Information Technology and the Good Life*. Springer US, Boston, MA, 687–692. https://doi.org/10.1007/1-4020-8095-6_45

[64] Alex S. Taylor, Siân Lindley, Tim Regan, David Sweeney, Vasillis Vlachokyriakos, Lillie Grainger, and Jessica Lingel. 2015. Data-in-Place: Thinking through the Relations Between Data and Community. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2863–2872. https://doi.org/10.1145/2702123.2702558

[65] Anna Vallgårda. 2009. *Computational Composites: Understanding the Materiality of Computational Technology*. Ph. D. Dissertation. IT University of Copenhagen, Innovative Communication.

[66] Anna Vallgårda, Laurens Boer, Vasiliki Tsaknaki, and Dag Svanæs. 2016. Material Programming: A Design Practice for Computational Composites. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM, Gothenburg Sweden, 1–10. https://doi.org/10.1145/2971485.2971554

[67] Anna Vallgårda and Johan Redström. 2007. Computational Composites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 513–522. https://doi.org/10.1145/1240624.1240706

[68] Rosa van Koningsbruggen, Hannes Waldschütz, and Eva Hornecker. 2022. What is Data? - Exploring the Meaning of Data in Data Physicalisation Teaching. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Daejeon, Republic of Korea) *(TEI '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 21 pages. https://doi.org/10.1145/3490149.3501319

[69] Gregory Vial. 2019. Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems* 28, 2 (2019), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003 SI: Review issue.

[70] Ron Wakkary, Sabrina Hauser, and Doenja Oogjes. 2019. Displacement: attending to the role of things in theories of practice through design research. In *Social Practices and Dynamic Non-Humans*. Springer, 151–171.

[71] Dakuo Wang, Josh Andres, Justin D. Weisz, Erick Oduor, and Casey Dugan. 2021. AutoDS: Towards Human-Centered Automation of Data Science. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 79, 12 pages. https://doi.org/10.1145/3411764.3445526

[72] Mikael Wiberg. 2014. Methodology for materiality: interaction design research through a material lens. *Personal and ubiquitous computing* 18, 3 (2014), 625–636.

[73] Mikael Wiberg. 2018. *The materiality of interaction: Notes on the materials of interaction design*. MIT Press, Cambridge.

[74] Robert K Yin. 2009. *Case study research: Design and methods*. Vol. 5. Sage.

[75] Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How Do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 22 (may 2020), 23 pages. https://doi.org/10.1145/3392826

Received January 2023; revised April 2023; accepted May 2023