Doing Data Ethnography: A Moderated Conversation and Reflection

Emma Garnett, Minna Ruckenstein, Tommaso Venturini, and Malte Ziewitz in conversation with Daniela van Geenen and Danny Lämmerhirt

Data practices abound to the point of becoming, what Marcel Mauss (2002, 100) has called, a "total social fact." They can be found in settings spanning from scientific to everyday contexts, public, professional and personal situations, ranging from counting and evaluating to compiling, commenting, complementing, translating, validating, ground truthing, contesting and otherwise modifying data. Such practices may serve mundane, economic, social, or political purposes. Moreover, they may enroll various artefacts, including online media and platforms, collaborative databases, crowdsourcing software, mobile apps, sensors, visualization and aerial imagery equipment, and all kinds of algorithmic protocols and procedures. The study and understanding of these practices merit empirical specification, in particular and as we will argue in this chapter, by ethnographic inquiry.

With the proliferation of digital technologies in practices of everyday life, scholars have expanded ethnography to accommodate, tackle, and account for the altered sites, settings, and situations of study. They developed and coined approaches such as "virtual ethnography" (e.g., Hine 2005), "online ethnography" (e.g., Markham 2005), and "digital ethnography" (e.g., Pink et al. 2016; Abidin and De Seta 2020). Such approaches increasingly recognized that studying the discursive-material dimensions of social interactions involving digital and online media requires examining the qualities of the medium facilitating and framing these interactions, as well (Marres and Weltevrede 2013; Marres 2017). Recently, scholars have started proposing ethnography in relation to digital data as an endeavor in its own right (e.g., Pink et al. 2016; Knox and Nafus 2018). These proposals explore similarities (e.g., Charles and Gherman 2019) and tensions between the fields of data science and ethnography, a relationship whose relevance and challenges the contributors to this chapter will also highlight. The development of *data ethnographic* approaches accounts for the observation that contemporary digital media practices are intertwined with diverse – more or less visible and understandable – data practices (cf. the current research program of the CRC Media of Cooperation (2020)).

How might ethnography engage with and attend to different data, the practices, settings, and infrastructures involved in their production and distribution? What methodological repertoires but also conducts could help us *do data ethnography*? How should data ethnography draw from, build upon, or expand existing methods in order to interrogate situated knowledges and (e)valuative practices that digital data are constituted by and are constitutive of? While some authors envision an ethnography of data as a way of "writing about society" (e.g., Lindgren 2020) by blending ethnographic strategies and computational methods, others emphasize the possibilities of ethnographic perspectives to question the principles and implications of producing knowledge through digital data (Knox and Nafus 2018). Hannah Knox and Dawn Nafus (2018)'s edited volume with the same title inquires into and carves out "ethnography for a data-saturated world," which the editors situate "at the interface of [the] two disciplinary traditions" of "quantitative [and] qualitative expertise" (24). Based on empirical studies, the contributors to the edited volume address what "ethnographies of data science" might look like, deal with the question of what it means to "know data," and demonstrate how experimenting with both data and ethnography can lead to new ways of knowing. This chapter adds to these discussions a more thorough understanding of the practice of *doing data ethnography*. It stages a moderated conversation between academic researchers with different kinds of ethnographic expertise and different levels of experience with Big Data technologies, data scientific training, and computational methods. By doing so, the chapter puts an emphasis on the relevance of reflection on the very notion of expertise in relation to (digital) data and their making.

In this book section, we propose "data ethnography" (cf. Knox and Nafus 2018, 24) as a situation-aware and -directed, flexible and expandable set of research strategies to explore data practices in situated ways. Ethnography starts with a commitment to observational fieldwork and "thick," multi-sited descriptions (Geertz 1983; Marcus 1995; 1998). Instead of (merely) seeing data ethnography as an expansion of anthropology by Big Data tools and computational methods (e.g., Laaksonen et al. 2017; Bornakke and Due 2018), we suggest data ethnography as a way of thinking and seeing differently how humans relate to the roles that (digital) data play in their everyday lives.

The endeavor and the title of this chapter were inspired by and allude to a recent special issue on "Doing Digital Ethnography," edited by Crystal Abidin and Gabriele de Seta (2020). The authors who contributed to this special issue were aiming at "laying bare their methodological failures, disciplinary posturing, and ethical dilemmas … acknowledg[ing] the messiness, open-endedness and coarseness of ethnographic research in-the-making." As ethnographic work is facing a wide variety of data practices, we need a methodological bricolage (Denzin and Lincoln 2005), methods that deal with messy and complex settings (Law 2004), and that use "inventive" (Lury and Wakeford 2012) strategies in tune with our research

interests, circumstances, distributions of expertise, and infrastructures. We, the editorial team of this chapter (Daniela van Geenen and Danny Lämmerhirt), chose the moderated conversation as a casual format to invite researchers to discuss their work and share their personal reflections with one another. The foundations for this conversation were laid at the fourth annual conference of Siegen's Collaborative Research Center *Media of Cooperation*, where our interlocutors discussed whether and how their own research could count as an ethnographic study of data practices. The goal of this discussion was to cast the web wide and to include researchers from diverse disciplinary backgrounds, with different conceptual lenses and methodological approaches to the study *of* and *with* data. With this setup we wanted to define and test the boundaries of data ethnography, as well as find commonalities and differences across research approaches. Our conversation partners are Emma Garnett (King's College London), Minna Ruckenstein (University of Helsinki), Tommaso Venturini (CNRS Centre for Internet and Society), and Malte Ziewitz (Cornell University). These scholars study diverging topics with different methodical approaches at the interface of ethnographic work and more digital entry points, ranging from conducting field work on sites relevant for inquiring into data practices to using computational methods and/or doing deep readings of datasets. Their subjects of study include interdisciplinary collaborations, the mundane work within a search engine optimization company, or the methodical opportunities of data sprints and digital methods.

The editorial team framed the discussion with open-ended questions in order to probe the term *data ethnography*, its methodological variants, and its possibilities for engaging with and examining data practices. Even though our conversation partners initially prepared their answers individually, the responses resonated surprisingly well with one another, and they helped the editorial team surface various common topics of ethnographic work *on* and *with* data, including questions of reflexivity and field construction, questions of how to establish collaborations in the field and also the shifting roles ethnographers play in the field. In response, the editorial team asked the conversation partners to elaborate on how the practice of writing ethnographically changes when focusing on data, on the possibilities and troubles of establishing required interdisciplinary and trans-institutional collaborations, and on how scholars may account for their own practices when doing data ethnography.

The conversation highlights that not everything is new when engaging ethnographically with digital data, but that data ethnographers can relate to and draw from a rich body of ethnographic studies and writing. Malte Ziewitz, for instance, addresses algorithmic profiling, or the traveling of medical records in clinical settings mostly through "traditional" observational methods of following things and people around. Minna Ruckenstein discusses that it is not the data practices per se, but the development of relevant research questions, that matters to her work. In

order for these questions to emerge, familiarity of the researcher with data practices can offer a useful starting point.

This draws our attention to the praxeological aspects of doing data ethnography. As our conversation will show, it is important to ask *how we do data ethnography* and how we can account for our own research practices and the methodological decisions we make when we are confronted and dealing with data. Such reflexivity can prompt important insights and questions about things we might otherwise take for granted: what we address when we use the term "digital data," the very nature of data practices, or how we can grasp and study their (perceived) everydayness.

For instance, our discussants address the vexing question of how to cope with distributed settings across scales. Instead of repeating the argument that ethnography has always dealt with large datasets, they argue that data ethnographies can devise methods to deal with distributed sites, for instance, by following people, physical records, or digital data around; by studying the properties of digital media as conditions for social interactions and relations; or by engaging with data visualizations. A reflection on ethnographic work on and understandings of data practices can also help to render debatable and reflect on professional data practices. As it has repeatedly been mentioned, in digital settings someone else's research methods might frame and structure one's own research practices (e.g., Lury and Wakeford 2012; Marres, 2017; Ruppert et al. 2013). Engaging ethnographically with digital settings can include turning professionals into co-interlocutors or establishing partnerships with "domain experts."

In conclusion of the moderated conversation, the collaborators reflect on the oxymoronic nature of the notion of data ethnography. On the one hand, we might dismiss the term as a fad of an academic landscape needing to come up with innovative terminology in order to receive research funding, for instance. On the other hand, our conversation partners underscore the usefulness of data ethnographic approaches to combine quantitative and qualitative methods in meaningful and fruitful ways and to open up spaces for dialog and reflection on disciplinary principles, differences, and possibilities of cross-fertilization. Perhaps the most notable common denominator in our discussion is a shared idea of the relevance of ethnographic sensibilities towards digital data and the practices they bring about. As our conversation partners emphasize, ethnographic approaches and the research conduct these approaches call for, have the advantage of creating surprising perspectives, unexpected questions, and, thus, new insights.

Generating New Insights from Data Practices

Daniela van Geenen and Danny Lämmerhirt: *A common distinction between data produced through ethnographic approaches and digital data builds on Clifford Geertz's famous distinction between "thick" and "thin" data (1973). For example, some scholars claim that Big Data or other (semi-)automatic information technologies produce "thin descriptions," whereas ethnographic approaches are said to produce "thick descriptions" (Bornakke, Due 2018; Laaksonen et al. 2017; Thompson 2019). Other scholars liken ethnography to Big Data as inductive research strategies revealing the meaning behind datasets (Charles and Gherman 2019). Your own work does not follow this prevalent but contested distinction (cf. Marcus 2011; Marres and Gerlitz 2016; Paßmann and Schubert 2020). Instead, your work renders visible how the datafication in/of society manifests in specific data practices – as an object of study – and in doing so, makes these practices interrogable and debatable. For example, you discuss how data can highlight new associations between people and their health beyond the individual, possibilities of doing issues with data, or materializing imperceptible things. What do you think is the role of ethnographic approaches in exploring new and unanticipated aspects of data practices, and what ethnographic sensibilities and methods could support this?*

Minna Ruckenstein: *In my own work, ethnographically oriented approaches guarantee that something new and unexpected can be found. In practical terms this means that the research process needs to be long-term and varied, to capture as many perspectives to data practices as possible. In addition to guiding the research, ethnographic sensibilities support participatory approaches that are important, if not necessary in this area. For instance, collaborating with the Berlin-based AlgorithmWatch (Ciusi et al. 2020) gave us the opportunity to be part of the interventionist public debate concerning the uses of automated-decision making and opened important conversations in this area (Ruckenstein and Trifuljesko forthcoming). Collaborations are risky in terms of customary academic practice, as they can position researchers as industry partners or activists. This kind of blurring of positions is, however, often a required move, because the expertise is located outside academia. I have found the anthropology of futures manifesto*[1] *that promotes a political and interventionist research stance, a good resource to think with. Data practices are contemporary worldmaking activities that call for engagements with the emerging and as yet unknown worlds of present, possible and desired futures. As ethnographic orientations are comfortable with open-endedness and messiness, they constitute an excellent support for the study of data practices.*

Emma Garnett: *Experience from previous research projects about air pollution has helped me anticipate how my methods and sensibilities might be able to contribute to the challenges different data practices raise for practitioners. Scientific understanding of air pollution is riddled with uncertainty, which means how it exists and gets recognized as a scientific, health and*

1 See https://futureanthropologies.net/2014/10/17/our-manifesto/.

legal concern is an ongoing challenge (see Murphy 2006). I have been interested in research design that allows these challenges to be engaged with. Participatory air pollution science is in some ways a response to these uncertainties because it attempts to include how lay understanding, or "popular epidemiology," might shape exposure risk (Brown 1997). Implicit in the design of research with portable and wearable sensing technologies is the often assumed relevance of data for those who participate in producing it. However, in the collaborations I have been involved in these devices frequently fail to engage participants or users in ways anticipated by researchers and designers. Rather than understand these encounters simply as a failure, ethnography has helped me reveal the differences between disciplinary approaches to data alongside those of the people who wear the sensors. I hope these exchanges can help open up space for debate about what kinds of data matter for understanding and responding to air pollution.

Tommaso Venturini: *Anyone who has ever worked with data (digital and non-digital, by the way) knows all too well that data is messy, dirty, full of unexpected sides – there is no better example of this than the air pollution case discussed by Emma. In most data projects, 90% of the time and effort is spent cleaning the data. And cleaning is the wrong word, because it suggests that one easily knows how to distinguish the noise from the information, which is rarely the case. "Data wrangling" is a better metaphor because it captures the struggle that is always required in order to extract the smallest finding from a dataset. But even the image of wrangling is inaccurate because it implies that the data analyst is in charge and knows exactly what she wants to obtain, which is seldom the case. Data practices are more often practices of reciprocal domestication in which a dataset and a data analyst slowly and sometimes painfully learn to live together through a mutual adaptation. The problem with the ethos of modern data science is that this process of reciprocal domestication is considered somehow shameful. Instead of being proud of the iterative and accommodating process through which dataset are cultivated and harvested, most analysts prefer to present their results as if they had always been present and manifest in the data – even through both their research questions and their datasets had completely transformed from the beginning to the end of their investigation. This positivist and objectifying vision of data practices is problematic for two reasons at least. First, because (as all forms of scientific and technical objectification) it conceals the many ways in which dominant groups use data to uphold and naturalize their domination. Second, because it hides many interesting things that happen in the data-accommodation and from which there is much to learn. Data ethnography, by drawing our attention to data as a process and as Minna puts it, "being comfortable with open-endedness and messiness," has therefore the great advantage of making visible and salient the network of reciprocal interferences that might otherwise be forgotten.*

Malte Ziewitz: *The possibility of wonder is one of the biggest strengths of ethnographic approaches – and if there is a field in which a bit more wonder would not hurt, it is the field of data practices and data. "Algorithms," "platforms," and "AI": the vast majority of categories*

we use to talk about these systems originates from marketers and engineers who had the privilege of coining them. While this should not come as a surprise to anyone (Gillespie 2010; Woolgar 1990), it can be difficult to free oneself from this conceptual baggage and challenge the assumptions, blind spots, and beliefs that come with them. It is all too easy to take data as a starting point and end up with a literature whose main concern is to make a system "fair," "transparent," and "accountable" – with little insight into the "data wrangling" (to use Tommaso's term) that is going on behind the scenes.

Ethnographic sensibilities can help us achieve at least some distance. By challenging ourselves to understand a problem from a different place, we can make the familiar strange and ask a different set of questions. Why, how, and by whom are these terms used? What is their currency for different members of the organization? What are alternative ways of accounting for the situation and how do these relate to the prevailing ones? A good example is a recent project, for which we accompanied a group of low-income people in Upstate New York to study how they cope with broken credit scores. One finding was that the intricacies of credit scoring algorithms were not the first thing on people's minds. Such a finding may sound trivial, but it has potentially far-reaching consequences. For example, initiatives that aim at making credit scoring more transparent may not be as helpful as it is often claimed.

Sites of Study in Data Ethnography

Daniela van Geenen and Danny Lämmerhirt: *Digital ethnographers have argued that digital, networked (and increasingly interactive) infrastructures complicate Marcus's (1995) idea of "multi-sited ethnography." Digital systems multiply and distribute the locations and situations in which data practices occur, and who is involved in these practices, as is also demonstrated by the sites of study that you frequent in your work. How does your work approach the distributed nature of data practices? For instance, how have you constructed the field, and scoped the settings and situations of data practices?*

Tommaso Venturini: *Actor-Network theory has a famous slogan: "follow the actors." It is a great way to remind the researchers that the subjects they study are neither passive nor motionless. They move, and act, and interfere all the time and more often than not across the convenient boundaries that disciplines, and scholarly approaches have built their investigation easier and more orderly. Asking the researcher to "follow the actors" is another way of reminding her of the distributed nature of all collective action.*

Following the actors is both easier and more difficult when digital technologies are involved (which is increasingly impossible to avoid these days). I have written a lot about how the traceability of digital technologies displace the cost of collecting records of collective actions from the researchers to media infrastructure, so I will not elaborate on this point (cf. Venturini et al. 2017; Venturini and Rogers 2019; Venturini et al. 2018). I'd rather highlight the way in which digital infrastructures make it more difficult to follow social actors because they

make it possible to move in a variety of new ways whose difference is hidden by digital convergence. This is a complicated sentence, but it means a simple thing: whereas a traditional ethnographer could count, at least to some extent, to the change of scene, or of material props, or characters when observing her subjects, a data ethnographer will consistently see the same thing: a bunch of people looking at a computer screen and typing on a keyboard. At a first-degree observation, all data practices look alike even when are widely different and this requires exerting special attention to what happens on the screen and through it.

Minna Ruckenstein: *A big part of the work is to understand what counts in terms of data practices, who is involved and affected, and how. With this information, we start to define the field. Here, conceptual work is also crucial. We might, for instance, think about how metaphors support the study of data practices. As an example, we discuss the metaphor of the "broken data" as a lens to explore data engagements (Pink et al. 2018). In another article, we develop the notion of situational objectivity for demonstrating how the framework of mechanical objectivity falls short when people translate physiological measurements to fit their expectations and everyday experiences (Pantzar and Ruckenstein 2017). The notion of mechanical objectivity (Daston and Galison 2007) suggests that data provides results that are accurate, consistent, dependable, and precise. In contrast, we argue that situational objectivity highlights the everyday as a domain of interpretation, reflection, and ambiguity, proposing an analytical entry point to data relations. Treating objectivity as situated underlines the fact that data encounters are not straight-forward and systematic, but tend to combine knowledge in a more eclectic manner.*

Malte Ziewitz: *Ethnography does not scale well. Leigh Star (1999, 383) made this point nicely when she wrote that "The labor-intensive and analysis-intensive craft of qualitative research, combined with a historical emphasis on single investigator studies, has never lent itself to ethnography of thousands." What can we do about it? Well, one response would be to enter a methodological arms race and ramp up our arsenal of ethnographic tools. For example, we could resort to multi-investigator studies with ethnographic teams (Neyland 2008, 122) or find salvation in technology and use some of the same devices we associate with data practices, such as video conferencing, bots, and self-reporting apps, blurring the boundaries between research and surveillance.*

The problem is that many of these approaches tend to assume that it would be desirable or even possible to come up with a more complete or comprehensive picture of a setting. Yet such scalar thinking goes against the grain of many of the sensibilities that appreciate so much about ethnography. Personally, I'd be more interested in understanding how notions of scale and scalability figure in particular settings – something that colleagues once called "scalography" (Woolgar et al. 2009). For example, whose ideas are we taking on when calling for an "ethnography of thousands"?

Against this backdrop, I do not think we need to be afraid of studying practices that appear to be distributed. Take, for instance, the case of what is commonly known as "web-based patient

feedback" in the British healthcare system, basically a form of online reviews for hospitals and Trusts. This work is happening in many different places, times, and systems all at once. Yet instead of striving to conduct an ethnography of thousands, it is equally plausible to just go for an ethnography of ones. In my own work, for example, I traced the journeys of specific postings from the patients' beds and living rooms through the feedback website's database and moderation system back into the wards and offices of hospitals and Trusts (Ziewitz 2017b). Doing so taught me a lot about the practices involved in managing the experiences of patients and making them travel through large organizations. It also showed me how the very idea of scale posed a problem not so much for frontline staff and patients, but for those who were in charge of turning postings into "data."

Emma Garnett: *Involving people (citizens, patients, school children) in air pollution research has created new spaces to do data ethnography and to contribute to data generation. I have already mentioned the dynamic of air pollution data practices in institutionalized science and public settings. These surface new challenges because different expertise, expectations and concerns come into contact. Accounting for the tension and incommensurability these encounters raise and specifying how, for example, promises of individualized interventions occupy a contradictory space with the social and environmental determinants of public health is something I wish to continue to explore. I have scoped settings and situations for staging differences between data practices and data in a documentable way (Fortun 2012): through organizing interdisciplinary data-analysis sessions; by contributing to creative participatory workshops; and by supporting the design of online spaces for sharing data practices, all of which involve thinking across a diversity of understandings and engagement with air pollution data. By laying multiple data side by side, these moments of dialogue produce what Mike Fortun, Kim Fortun and George Marcus (2017, 19) describe as "kaleidoscopic logics," in which different data can be leveraged to produce unexpected insights and allow for "explanatory pluralism."*

Daniela van Geenen and Danny Lämmerhirt: *Your responses refer to ethnography and ethnomethodology, and how both research traditions might relate to the study of digital infrastructures and the data practices they facilitate and produce. In Garfinkel's ethnomethodological tradition (1967) methods are understood as the everyday ways, in which the members of a specific work setting develop specific professional routines, or in which people organize and understand their lives collaboratively. "Following the actors" here becomes a research activity that attempts to stick as closely as possible to the terms of these members of a specific setting. In all of your responses there is a – more or less explicit and renewed – focus on the idea of following (social) actors, or becoming familiar with the sites and environments of study. Moreover, in light of data practices the need of doing efforts to "domesticize" data is emphasized. What are the primary challenges in following practices related to digital data, technologies and infrastructures? To which extent do you as a researcher attempt to "follow the actors" whose practices you study including their understanding of the conventions (i.e. terms) that frame and inform their practices?*

Tommaso Venturini: *I think that Malte in his previous answer puts the finger on the problem: how do we accommodate a set of research techniques that were originally developed to study communities living on exotic islands or in the deep of the forest, to study the practices mediated by digital infrastructures? How do we scale up ethnography, but also how do we make it more distributed? The problem is not only that we are investigating phenomena that mobilize more actors (though this is certainly true), but also and crucially that these actors are dispersed geographically, socially, and culturally. Classic ethnography can count on a certain unity of action: the interactions that it observes take place in a situation that is well-defined and well-delimited. But what does it mean to do situated research, when the actions we observe are distributed across different continents, languages, cultures, social contexts? What does this extreme multi-sited ethnography look like?*

Emma Garnett: *The primary challenges I have encountered in following the practices of digital data, monitoring and sensing technologies and their wider knowledge infrastructures is that they are hard to identify or distinguish. They often look the same, as Tommaso highlights. Largely, I seem to have addressed this by focusing on different kinds of work and labor: from sitting alongside my collaborators at computer screens, to contributing to collated data sets in shared spreadsheets, to moving sensors around a city to facilitate "continuous" data collection. This has become easier recently because I have had the opportunity to learn how to implement the smaller and cheaper technologies increasingly in research and engagement contexts. By getting to know the devices I am able to experience their errors/challenges and therefore work around problems in practice with collaborators. The limits of data ethnography can also become overwhelming in these moments when, for instance, the questions my ethnographic insights raise are incommensurable with the epistemic commitments of scientific or health data. It is exciting to see work that is starting to interface qualitative and quantitative data through novel forms of digital analysis (see Blok et al. 2017). For example, Kayla Schulte and Karl Dudman have developed an open-map of "air quality anecdotes" – shared and uploaded by citizens in the city of Oxford to complement sensor data generated in the same geographic area*[2]*. In terms of the extent to which I include the understanding and conventions of the practitioners I "follow," one term I have been working with is "person-centered environments," which was coined by my computer scientist colleagues. By critically engaging with the terms deployed in their practices I have tried out different ways of working through different ways of understanding the problem of air pollution. This has allowed for the inclusion of how people participate in and negotiate science-led data practices, which can be used to encourage researchers to account for the social and cultural relations and structural factors that are reified in a data point.*

Malte Ziewitz: *These are really good questions, and I don't think I have an answer. One strategy for dealing with these problems would be to pause and reconsider just how ideas like "well-*

2 https://viewer.mapme.com/oxford-airquality-openmap.

defined," "well-delimited," and "multi-sited" have achieved such currency in contemporary debates about data practices and data. For example, to what extent do our ideas of data practices depend on and perpetuate an ideology of openness and scalability that has worked wonders for the early theorists of cyberculture and the entrepreneurs and venture capitalists who followed them (Turner 2010)? Is it possible that our difficulties with fielding data practices are a problem of us taking on a set of corporate metaphors like "platforms" that so elegantly combine a sense of being indispensable with an abrogation of responsibility? One exciting prospect here is that we have an opportunity to rethink some of our own strategies and preconceptions as ethnographers and try out different things. The digital methods initiatives have shown that this is possible and theoretically generative, as have more conventional approaches that did not follow human actors, but data, categories, and algorithms.

Doing Data Ethnography and Writing Ethnographically

Daniela van Geenen and Danny Lämmerhirt: *Based on your own work, through what approaches have you studied data practices? How did your study of particular data practices attune methodologically to different types of data, data practices and devices?*

Minna Ruckenstein: *The research projects that I have led and collaborated with in this area cover a wide range of data practices, each calling for specific methodological approaches and arrangements. A lot of the work that I do is fairly conventional ethnographically oriented qualitative research. Asking people what they do with the data and what kind of value it has for them, is always a good starting point. Overall, I am more interested in thinking about research questions than data and devices. Questions that are worth pursuing tend to emerge after you have familiarized yourself with the data practices. Digital data offers the possibility to see things in a way that would be impossible otherwise, but that means that one needs to do research on, or with digital data, reflect on how the data is used, by whom, and for what purposes, and then depart from the customary uses. We have, for instance, studied the collective rhythms of the heartbeat (Pantzar, Ruckenstein, and Mustonen 2017), and the patterned nature of individual experiences with antidepressants (Ruckenstein 2019). Digital data can be aggregated to identify collective patterns that have to do with health, everyday mobilities, time use and environmental exposure (Nafus 2019). These kinds of studies require unconventional uses of digital data, but the exploratory aspect also makes them particularly exciting.*

Malte Ziewitz: *My own approach to studying data practices has not been fundamentally different from the one I use to study any other practice or phenomenon. For me, the beauty of ethnography lies in its situatedness and the need to navigate a setting in a way that is uniquely adequate for the circumstances of inquiry. What matters most, then, is not any particular tool or method, but the movement we engage in as ethnographers. Becoming an insider while being an outsider, and using the contrast to highlight what is taken for granted in and*

about the setting – I think this motto, which my supervisor taught me when I was a graduate student, still holds for much of what I do today.
Over the past few years, I have used this way of thinking to explore a number of settings in which data or data practices are salient. On one occasion, for example, I was curious about the figure of the algorithm and came up with the idea of going on an algorithmic walk. We went out, came up with an ad hoc algorithm to provide directions, recorded our observations, and reflected on the work it took to put that algorithm into practice (Ziewitz 2017a). On another occasion, we teamed up with people who were trying to repair a broken credit score and accompanied them through a mix of diaries and interviews over the course of an entire year (Ziewitz and Singh 2021). As different as these cases were, the underlying premise was the same: what can we learn about a setting from engaging with its members on their own terms?

Tommaso Venturini: *It is difficult to give a short answer to this question, because data practices are so varied and so fundamentally diverse that most of the time it is necessary to develop ad hoc protocols to study them. The one thing that is crucial, however, is to allow oneself the time to become deeply acquainted with these practices. This is a classic ethnographic piece of advice: hang out as long as possible in the situation that you are studying, stay with them and with their troubles until you feel the risk of "going native" or as Malte beautifully put it "becoming an insider while being an outsider." This is true for the study of data practices as for the study of any other practice – I completely agree with Malte that the basic ethnographic movement remains the same. This is not to rule out "quick and dirty" approaches, which I am a big fan of. Yet, while these approaches can be extremely useful when doing research with digital data, they are generally insufficient when doing research on digital data and on the infrastructures and situations that uphold their creation.*

Emma Garnett: *Much of my ethnographic research has been conducted in interdisciplinary projects about air pollution. These collaborations have largely been methodological in focus, responding to the persistent challenge of how to generate air pollution data in ways that are relevant for public health, policy and publics. It was during my PhD that I came to realize studying data practices animated the social dynamics of interdisciplinary scientific research and materialized the unspoken aspects of knowing. For example, by paying attention to how scientists use technologies to cultivate an intuition for air pollution in research, I could describe the embodied and imaginative work that goes into making data (Myers 2015). This enabled me to attend to different data practices and types of data – monitored, modelled, statistical – and understand why working out what counts as "good data" for my collaborators was so tricky (Garnett 2016; 2017a). Studying data practices in interdisciplinary science has shaped my ethnographic research since, particularly in response to portable and wearable sensing devices that are extending and distributing data practices to a range of publics. Rapid changes to the kinds of data and data infrastructures required to accommodate these informs how I approach fieldwork, which I try to situate at the various intersections of different data and data practices.*

Daniela van Geenen and Danny Lämmerhirt: *Scholars have often discussed the role of new technologies to afford new ways of writing ethnographically, for instance, by building on digital logs in online communities, or making use of data visualizations. Likewise, you all engage with data and digital technologies in interesting novel or uncommon ways to produce, document, and write new insights about data practices. Can you elaborate on the role data and data practices play in your work for writing ethnographically?*

Minna Ruckenstein: *New modes of ethnographic writing are needed when data and data practices are included in the ethnographic mix. Visual methods can bring into productive tension what is often treated as opposites: objectifying life with metrics and carefully navigating social worlds by means of ethnographic inquiry. We were moving to this direction in our research on self-tracking, but we did not get nearly as far as we would have liked to. We followed a participatory research design that shares features with the "ethno-mining" (Anderson et al. 2009), combining the collection and analysis of quantitative data with qualitative data in an iterative framework. We documented how self-tracking transforms physiologies into information and feeds it back to people in visual format, promoting and intensifying sensory and informational attachments. By following how people discussed their newly visible physiologies, we could demonstrate how, once visualized, the data triggers new kinds of ties between people and their measured actions and reactions (Ruckenstein 2014). What we were not able to do as well, however, was to take advantage of the rich material that we had, in telling other kinds of visual stories with the data.*

Tommaso Venturini: *Writing ethnographically, to me, is to be able to render the situation in which the research findings have been produced, to transport the reader in the presence of the subjects that are described and make her able to hear their voices and feel their presence. This can of course be greatly helped by digital technologies, but more through their multimedia affordances than strictly through data.*
Writing with data – i.e. producing documents (scientific articles or reports) that encompass not only text but also numerical, categorical or relational data – is extremely interesting, but also very challenging, as Minna points out, because it require an attitude that is radically different from the traditional way of writing of qualitative social sciences. On this specific skill, I believe that the tradition of vivid ethnographic writing (with its stock of illustrative vignettes and quotes from informants and field notes) should mingle with the capacity to illustrate findings through diagrams, charts and equations typical of natural and data sciences. In natural and data sciences papers, the text is there to guide the attention of the reader, provide additional information and contextualize the findings, but it is the figures that make the argument. Most social scientists (including myself) are still miles away from the sophistication that allows some natural scientists to deliver complex messages through visual languages. If we want to conduct data ethnography properly – that is, taking into account the ethnomethods of the communities we study – we should learn to master this type of writing too. This can help unpacking the meaning that our subjects of study infuse in their figures and equations

(as Emma rightly points out), but also teach us to render our own observations in a style that is not completely disconnected from that of our informants.

Malte Ziewitz: *I do not think I have been particularly innovative in this regard. Personally, I am a big fan of some of the more interactive formats people have experimented with. Laura Watts and Dawn Nafus's (2013) Data Stories come to mind; Kate Crawford and Vladan Joler's (2018) essay Anatomy of an AI System with its zoomable diagram; or the pleasantly unwieldy Asthma Files assembled by Mike and Kim Fortun (2013) and their team. An early piece I really loved was Bruno Latour and Emilie Hermant's (2006) Paris, Invisible City. Of course, such experimentation will not always be successful. But we don't know if we don't try.*
In my own work, I have been interested in writing as a method of inquiry. As Bruno Latour puts it, "if science is a practice, and social science is a practice, then we need to know what sort of practice writing is" (Blok and Jensen 2011, 163). We have a wonderful group here at Cornell called Historians Are Writers! (or HAW!) that has been dedicated to this work.[3] *It is led by historian Aaron Sachs and his students, and fortunately open to ethnographers, too. Among other things, we read a range of texts from novels to nonfiction essays to learn from them for our own writing. How do you open a chapter? How do you write about others? How do you speculate well? Generally speaking, I find that writing is most powerful when it manages to exemplify its point through all the tools we have available as writers, including form, tone, voice, and drama. Learning to write artfully is critical for our work, but also strangely underappreciated.*
So what does all this have to do with data? I do not think there is a fundamental difference in writing ethnographically about data practices. I'd say more generally that we would do well to be more courageous and experiment with forms of writing that are uniquely adequate to our topics and ideas.

Emma Garnett: *When interviewing atmospheric chemists, mathematicians or epidemiologists about their data practices I am often shown illustrations of how air pollution is calculated. On one occasion when I was inquiring into how statisticians check data sets for error, a colleague jotted down an equation to show me what specific aspect of the calculation they are testing to determine the temporal effects of air pollution on population health. In interdisciplinary meetings, researchers often center discussions on data – as objects of shared interest – represented or visualized in ways that highlight specific insights or omissions. By carefully unpacking how different methods, tools and analyses fostered ways of seeing and investigating air pollution (Coopmans, Vertesi, Lynch, and Woolgar 2014), I was able to write about air pollution data practices from different perspectives and ontological starting points. However, it also made me realize the limits of ethnographic writing. I am less adept at finding creative ways of sharing my findings, beyond re-presenting the data visualizations produced by*

3 Historians Are Writers! (n.d.) Available at: https://historiansarewriters.wordpress.com/ (accessed 29 June 2020).

the researchers I study or providing a written analysis of a creative or sensory practice. This feels dissatisfying and I am learning from other fields and colleagues who do this in ways far better than I have. Adding to Malte's great examples, there is also the xcol's Ethnographic Inventory[4] *which shares tools and resources for ethnographers to take better care of their relationships with data and data practices, including how to write collectively about them.*

Daniela van Geenen and Danny Lämmerhirt: *How do you define and account for the data practices of your research participants/subjects and your own data practices?*

Minna Ruckenstein: *I approach this question from the larger perspective of datafication, signaling a broader trend across societal domains, in fields from health and media to public administration, in political life and the private sphere. While the tracking and surveillance of everyday actions of consumers and citizens is expanding and becoming ever more fine-grained, the everyday gets tangled up with data practices. We – including research participants and myself – contribute to data practices when we purchase goods and services online, engage in self-tracking practices, visit a medical doctor, take part in customer loyalty programs, use online search engines, or upload content to social media platforms. Data practices should be seen as everyday practices, otherwise we lose sight of the infrastructural changes that are currently shaping ourselves, collective formations, organizations and societies. One of the questions that I am interested in is how the market "sees" the consumer with the aid of data and devices – and how people react to that seeing and modify their behavior accordingly (Ruckenstein and Granroth 2020). The use of data and rules for calculation and prediction have a much longer history, but a shift can be detected in the way the market operates as a classifier: personal records and the scores and segments derived from them are now tradable objects that act back on people, shaping intimate experiences and defining modes of sociality (Fourcade and Healy 2017).*

Emma Garnett: *How I define and account for my data practices and the data practices of those I research and work with has also evolved over time. In my PhD I mainly observed data practices and it was through this work that I learned about the capacity of data to bridge different understandings of air pollution. By learning about the ways atmospheric chemist's modelling of air pollution changed in response to the epistemic requirements of epidemiologists interested in a proxy measure of human exposure, I began to think about how ethnography might participate in the multiplicity and differences of air pollution research in more productive and creative ways (Garnett 2017). I have built on this insight in my ethnographic fieldwork through which I try out different ways of working with data to create dialogue with my collaborators. This also requires finding new ways of accounting for my own data practices, for instance by producing qualitative data of social experiences of air pollution alongside plotted data points of exposure. In 2017 I had the opportunity to work with architect and*

4 https://xcol.org/ethnographic-inventory/.

researcher Nerea Calvillo, which helped me learn about a different way of doing data ethnography. The Yellow Dust Sensing Infrastructure by C+ Arquitectas translates data into a mist which means levels of air pollution can be experienced in an embodied way (Calvillo 2018; Calvillo and Garnett 2019). As an ethnographer examining the design of the infrastructure and how people and different publics interacted with it, I realized that finding creative ways to extend participation with air pollution challenged common ways of approaching data. We, as researchers, became subject to the meanings' visitors attached to the installation through their engagement with data which influenced our own data practices.

Malte Ziewitz: *What makes the idea of data practices so interesting is that both we and our interlocutors are continuously engaged in them. Howard Schwartz (2002) made this point in an essay he provocatively titled "Data: Who Needs It?" As a professional group, he argues, social scientists are interested in "finding and imposing normative standards on the collection, display, and use of data" (Schwartz 2002, 7). The problem is, of course, that our interlocutors do very much the same. Just think of Emma and her work with people monitoring air pollution – two interlocking and partially overlapping inquiries into data practices and data. If we look at it this way, one methodological challenge we are facing as ethnographers consists in managing a tension. On the one hand, we cannot single-handedly impose our own methodological assumptions on the setting without losing much of what makes it special. On the other hand, we cannot just "go native" and abandon our analytic projects by taking on our subjects' methodologies. In the case of data practices, the situation is even more confusing because the practical and intellectual labors of our interlocutors tend to resemble our own – or are entirely indistinguishable. So how to deal with this predicament?*
One way of managing the tension would be to settle on the ethnographic formula I mentioned earlier – and Tommaso subscribed to, as well: becoming an insider while being an outsider, and using the contrast to highlight what is taken for granted in and about the setting. A complementary approach would be to consider interlocutors not as research subjects but as epistemic partners and engage in what anthropologists and STS scholars have called para-ethnography, i.e. "an analytical relationship in which we and our subjects – keenly reflexive subjects – can experiment collaboratively with the conventions of ethnographic inquiry" (Holmes and Marcus 2008, 596; see also Fortun 2001; Dumit 2004). In my own work, I have found these sensibilities incredibly productive. Rather than studying search engines, for instance, we can study (with) those who study them: website owners, search marketers, users, engineers, and even regulators (Ziewitz 2019).

Tommaso Venturini: *Digital methods are, in many ways, an update of ethnomethodology. Exactly as Garfinkel has taught us not to constrain the interpretation of social situations within the external categories of academic theoretical framework, but to learn from the methods that social actors themselves employ to make sense of their own worlds, so the idea of "the methods of the media" (as Richard Rogers calls them) is to stick as closely as possible to the way in which the actors of digital spaces act and describe their actions (2013). Online platforms,*

digital companies, public administrations, Quantified Self adepts and basically everyone who is dealing with data is, in one way or the other, involved in the business of doing social research – even if that happens outside the walls of academia. If there is something that characterizes our contemporary societies, it is that we are all constantly dealing with practices of quantification and social researchers have to learn to live with that. As Minna puts it, datafication is everywhere, not just in academia, and we cannot but reckon with it. We, social researchers, have never had the monopoly of data production, analysis, or visualization, but now we cannot but recognize that we are at the fringe of the data industry and that, marginal as we are, we are forced to search for alliances. This can be done by practicing the kind of para-ethnography suggested by Malte, but also by doing all sorts of para-data-science where we collaborate with data practitioners outside academia to appropriate and repurpose their datasets and computational tools.

Daniela van Geenen and Danny Lämmerhirt: *Your answers point to quite different roles that ethnographers can play when dealing with data, related technologies, and practices. Instead of confining ethnography to observation, ethnographers can play the role of co-investigator to render data practices strange, or interventionist to break with dominant methods. This has implications for how methods co-create the field and are part of the field (Law 2004; Lury and Wakeford 2012). It also demonstrates that collaborations or "alliances" can be imagined in various ways that exceed organizational setups and can be part of one's approach to studying data practices. What has influenced your choice of role during your research and how you have reflected on them as part of your studies, especially in relation to ideas of collaborations and "alliance-building"? And what could other ethnographers who seek to collaborate learn from your experiences?*

Tommaso Venturini: *Working on digital social data we have an advantage that is precluded by most other data ethnographers: these practices are so newly established and so poorly consolidated that there is space for us to take the role that we want. While studying the practices of older and more established data initiatives (a medical records archive, for example), you are immediately pushed to the role of observer. The data are too precious, specialized and protected for you to touch them without the risk of polluting them. This is not the case for digital data about social phenomena. None is sure what to do with them anyway, so you are generally welcome to suggest your ideas, try your tricks, and manipulate the records. There is little risk that you break anything, because not much has been built yet. This openness is probably temporary (we have seen that platforms APIs, for instance, are already closing down) but, while it lasts, it's a great occasion to do participant observation and not just observation.*

Minna Ruckenstein: *Tommaso is speaking from the position of an expert data ethnographer in a sense that he has the technical skills to do data work. Since I have no coding skills, or even skills to use most digital methods, I need to collaborate with others to get things done. This means that in projects that involve digital methods, and the exploration relies on data ana-*

lytics, I am not fully in charge of the workflow. This has been a humbling experience in many ways, highlighting my marginality in relation to the questions that I am studying. On the other hand, this is also a good position to be in, because it underlines the exclusive nature of methodological expertise. Feeling unskilled probably explains why I have been drawn to uncover hidden competencies and knowledge in relation to digital data. In our Big Data project, I ended up working closely with content moderators, who had shaped the Suomi24-data that we were using, with their removal decisions. The perspective of experienced moderators opened an "emic" view on the everyday data work and the conversational dynamics of the platform (Ruckenstein and Turunen 2020), opening new venues for thinking about who participates in the data production.

Malte Ziewitz: *In my experience, our choice of roles as ethnographers is rather limited and usually beyond our control. Like Minna, I have not been trained in data science and would not pass as a data science expert by any of my colleagues' standards. However, that has not prevented interlocutors from asking me to help them analyze a data set or conduct statistical analyses they thought I should be able to pull off. Others, in contrast, have felt obliged to tell me what a URL is and would not trust me with a screenshot. I've found it helpful to think about those roles as data that can tell me something interesting about the field. In fact, this back-and-forth appears to be a crucial element of building these alliances. That said, I'd like to point out that antagonism can be an equally important strategy. Alliances are all fine, but when you're studying data practices associated with Big Tech, these collaborations can be difficult to build. Even worse, we've seen them fire back or end up being criticized as treacherous.*[5] *I like the work of Carl DiSalvo (2012) on adversarial design in this regard, an investigative practice that uses the means and forms of design to challenge beliefs, values, and what is taken to be fact. Maybe it's time to cultivate "adversarial ethnography" as an alternative.*

Emma Garnett: *Many of my ethnographic roles in research collaborations have been shaped by funding calls and institutional structures. They are determined by disciplinary and career hierarchies in ways that can be challenging. These experiences have led me to consider your question on a few occasions, in terms of how to develop collaborations or alliances that can both facilitate ethnographic study as an early career researcher and ensure it is interesting and valuable to the people I work with. In the project about personal sensing technologies, my aim has been to explore and demonstrate the value of doing data ethnography to science and policy. This has taken time because it has involved seeking out collaborators to work on a topic that often excludes social science knowledge and understanding. It has also required accepting things not working or going to plan and developing ways of doing data ethnography differently. For instance, observing the participatory aspects of science by attending users*

5 A good example is the so-called emotional manipulation study conducted by Facebook and – among others – Cornell University researchers, which has been widely criticized on ethical grounds (see, e.g., Felten 2014; Flick 2016).

doing personal monitoring was not always possible in the projects I worked with. On these occasions, I worked with social science colleagues to conduct interviews or focus groups with different groups– patients, citizens, activists, researchers – and then analyzed the transcripts with computer scientists, public health researchers, and participants in relation to the digital maps of exposure and health data. I also started to build relationships with other groups working in the field of air quality to test what I came to learn and to see whether I might be able to develop other approaches, for instance by multiplying and demarcating the field again or by including new approaches and expertise within it.

Daniela van Geenen and Danny Lämmerhirt: *When reading your responses, you also highlight ethnography's potential to explore and understand data as constructed in and related to social situations and settings. All of your contributions highlight data practices as an entry point to explore, understand, and reflect on the broader context of the datafication of society and, thus, the significance of digital infrastructures for everyday life and practices. To give two examples, Emma says that data ethnography's opportunities to add sociological perspectives on data, in particular in institutionalized knowledge production settings such as the measurement and monitoring of environmental issues (e.g., air quality) in order to promote and facilitate science that is not just "effective" but also aware of possible ethical issues and challenges. Minna, on the other hand, denotes data practices as "contemporary worldmaking activities" that provide the researcher with opportunities to approach and pay attention to the social embeddedness and meaning of data. Can you elaborate on how ethnographic approaches can help augment social understandings of digital data, and whose understanding this concerns? Are there lessons to be learned to not just help promote specific, or unilateral (social) understandings of data (e.g., those of dominant groups)?*

Minna Ruckenstein: *While a large digital dataset can offer indispensable support for understanding the scale of the studied online phenomenon, the data resources remain detached from the activities that precede data production. Ethnographic approaches are needed for recuperating the fact that a digital dataset remains insensitive to the intensity of circumstances and actions of those who generate the data traces. Since human experience and activities are inseparable from data, ultimately nothing that happens online is irrelevant to the data traces left behind (Pink and Lanzeni 2018). It is as part of messy and inconsequent online lives that Big Data about everyday lives is gradually accumulated. The same goes for data work, or data labor. The involvement of humans needs to be repeatedly reminded of, as it tends to disappear from sight. It becomes "ghost work" (Gray and Suri 2019). We are currently looking at a case of prisoners training AI for a Finnish AI company (Lehtiniemi and Ruckenstein 2022). Here, we use prison data labor as a starting point for exploring the collaborations, frictions, and power relations around data-based automation. Prison data labor offers a fringe-case for thinking about questions related to human data labor. Here, the adversarial ethnography, suggested by Malte, focuses on uncovering relations that might not become visible in any other context.*

Emma Garnett: *One way in which I think ethnographic approaches can help augment social understanding of digital data is by describing why different data cannot be pieced together like a puzzle to gain a fuller understanding of phenomena. This is a perennial issue for social scientists working in contexts like healthcare where their contributions are often framed as providing the context or lay perspective to an a priori matter of fact. Data ethnography is helpful because it accepts data and data practices are always situated, an insight that can be helpful for those only involved in one part of data work, such as the analysis. In recent work with colleagues in the Digital Humanities and Environmental Research Group at King's College London, we have been engaging with the "problem" of citizen generated air quality data for science and policy. It is a concern that has largely been addressed technologically, by testing the quality of data different commercial sensors produce. Building on the concept of "just good enough data" (Gabrys et al. 2016) we have attempted to subvert the knowledge deficit model by detailing why evaluating data on scientific terms alone fails to account for the knowledge, insights and actions enabled through citizen-led data practices.*

Daniela van Geenen and Danny Lämmerhirt: *What would you say are important data ethnographic sensibilities and how does one develop them? What could they contribute to future research as well as the broader societal debate around datafication and data-intense infrastructures?*

Tommaso Venturini: *Ethnography is an "internal research method"; internal in the sense in which people speak about "internal martial arts." It's force comes not from an external equipment but from its capacity to train scholars in a very particular discipline of attention, that allow them to notice minute things that are lost to a more casual observation (what Anna Tsing calls "the art of noticing" (2015)). This is the kind of sensibility that we need to develop for data practices as well: the art of noticing the subtle shifts that allow traces to turn into records; records into data; data into findings; and findings into evidence.*

Minna Ruckenstein: *As all of us have brought to the fore in this discussion, ethnography is a powerful tool for challenging the invisible dynamics, processes and power in complex socio-technical systems. An important role for data ethnography is in committing to concerns that are currently neglected, and in bridging elements that appear as unbridgeable. Ethnographic inquiry reminds us that our relations with technology are not only functional but also moral. Machines fail to care about real-life consequences, and this is something that we should keep in mind.*

Emma Garnett: *As we have all detailed in this discussion, data ethnography is a continuation of ethnographic sensibilities writ large. During my research I have taken on different roles and positions that are shaped by the particular circumstances of collaborations. When reflecting on how to specify data ethnographic sensibilities I was reminded of recent debates about contemporary ethnography that engage reflexively with its practices and methods. Adolfo Es-*

talella and Tomás Sánchez Criado's (2018) edited book Experimental Collaborations pays attention to different "fieldwork devices" that unfold when anthropologists engage with counterparts in the field who are also members of epistemic communities. Their experimental and inventive approach is helpful for characterizing data ethnography, because it implies working out ways of making data researchable with others. Finally, as Minna writes, data ethnography is about the building of social worlds rather than only observing them which means it is an approach that can also benefit from sensibilities developed in related fields, in particular feminist and intersectional approaches to data (D'Ignazio and Klein 2020).

Defining Data Ethnography

Daniela van Geenen and Danny Lämmerhirt: *How would you define data ethnography? Would you argue that there is a specific value/benefit in coining ethnographic work on/with digital data data ethnography, and if so, why and how?*

Tommaso Venturini: *The notion of data ethnography is an intriguing oxymoron. I've always felt torn, as a researcher, between my fondness for ethnography and my interest in data – a bit like a child asked to choose between mum and dad. Because ethnography and data, there is no point in denying it, do not go well together. Of all the techniques of social research, ethnography is the one that cherishes the most the unmediated exposition to the messiness of collective situations and the exploration of the thick networks of meaning and interpretation that twists and turns the smallest action. Data, quite the contrary, brings with it the promise of a liberation from the details of particular social situations and the possibility to extend one's view to embrace patterns and trends spreading far away in time and space. Of course, both of these visions are idealized. Ethnographers have always had to consider the way in which their subjects of study overflow the here and now in which they are observed, and social data are always messy and connected to the specific conditions of their production. But the tension between the two remains and explains why most researchers feel more comfortable in choosing to work with either qualitative or qualitative methods, or even to adopt a "mixed methods" approach that juxtaposes but does not really mix the two. Data ethnography is to me very close to the ideal of quali-quantitative methods that I have been pursuing (without ever achieving of course) in my whole career – a research capable of following networks of actions that expand far in space and time, but without as little as possible of simplification of aggregation.*

Minna Ruckenstein: *There are many ways of defining data ethnography and positioning ethnography in relation to digital data: ethnography of data, ethnography with data, or ethnography as data. Each of these offers a different perspective to the topic at hand. Our research mostly engages with the ethnography of data, and studies data and data practices as an object of ethnographic analysis and critical inquiry. Doing ethnography with digital data*

requires collaboration with data scientists, and careful mixing of qualitative and quantitative analytics. I am very intrigued by this kind of work, as it promotes an experimental and open-ended research stance. Yet, it is much more laborious, because it requires interdisciplinary collaboration and careful negotiation of shared aims. A further option for defining data ethnography is to emphasize data-awareness that translates into an attentiveness to the transforming effects of data practices in people's lives, but is also defined by a readiness to use digital data for uncovering the patterned nature of everyday phenomena. A powerful aspect of the data generated by self-tracking devices, for instance, is the possibility to transcend and bypass familiar ways of approaching bodies and lives: data becomes a resource for raising new kinds of questions and perspectives for inspection. By actively using data streams as part of research designs, ethnographers can contrast and move between human and machine categorizations of life, experiment with quantitative and qualitative approaches, and overcome the preset and the normative in order to learn something new.

Emma Garnett: *Based on my responses to your thoughtful questions, data ethnography might include approaching data as processual social and material forms with an attentiveness to the opening up of data practices to further interrogation and experimentation. Data ethnography will always be shaped by the identity, experience, and concerns of the researcher and those they work with. For instance, I don't have a background in digital methods or computer science but, as Dawn Nafus (2018) has described, my methodological experience and training means ethnographic sensibilities have affected the design of some of the data driven research and analysis I have been involved in. Perhaps it is this aspect of data ethnography that indicates the importance of defining data ethnography in ways that mean its different contribution to understanding and engaging with data and data practices can be valued and harnessed. In my research I am often pushed to explain how my interest in scientific data practices speak to more narrative or social accounts of data-shaped concerns in everyday life (Pink et al. 2018), so it tends to be something I negotiate and work out with others.*

Malte Ziewitz: *As your question makes clear, the value of a definition will depend on the specific circumstances of its use. So, a sarcastic answer would be: let's make it whatever helps us get the grant and pacify that notorious Reviewer 2. At the same time, I do believe that there is value in resisting definitions. Ambiguity, for instance, can be surprisingly productive. I learned this lesson when I co-organized a conference on the topic of "Governing Algorithms" at New York University back in 2013. In preparing the event, we easily could have defined the disciplinary boundaries of the event by resorting to a textbook definition of the term. Instead, we kept it reasonably open (Barocas et al. 2013), and a lot of people found this unsettling at the time. Strikingly, however, the ambiguity did not prevent them from engaging with each other across a range of disciplines, including media studies, STS, computer science, sociology, anthropology, and law.*

The point here is that in the same way we can take advantage of the indexicality of language to push specific definitions, we can also make strategic use of ambiguity to generate

new thoughts. Misunderstandings can be fantastic. In the best case, they make us stumble over our own ideas and engage us in a shared project of inquiry. So, I am quite happy with resisting your request. That said, the definition you suggest is a good starting point for us to play with the idea.

Daniela van Geenen and Danny Lämmerhirt: *Thank you so much for the conversation.*

This conversation was conducted as part of projects at the DFG-funded Graduate School 1769 "Locating Media" and the DFG-funded Collaborative Research Centers SFB 1187 "Media of Cooperation" – Project-ID 262513311 – and SFB 1472 "Transformations of the Popular".

References

Abidin, Crystal, and Gabriele de Seta. 2020. "Doing Digital Ethnography: Private Messages from the Field." *Journal of Digital Social Research* 2 (1): 1–19.

Anderson, Ken, Dawn Nafus, Tye Rattenbury, and Ryan Aipperspach. 2009. "Numbers Have Qualities Too: Experiences with Ethno-Mining." *Ethnographic Praxis in Industry Conference Proceedings* 2009 (1): 123–40. Accessed April 17, 2021. https://doi.org/10.1111/j.1559-8918.2009.tb00133.x.

Barocas, Solon, Sophie Hood, and Malte Ziewitz. 2013. "Governing Algorithms: A Provocation Piece." SSRN Scholarly Paper for the "Governing Algorithms" Conference at New York University. Accessed April 17, 2021. https://doi.org/10.2139/ssrn.2245322.

Blok, Anders, Hjalmar B Carlsen, Tobias B Jørgensen, Mette M Madsen, Snorre Ralund, and Morten A Pedersen. 2017. "Stitching Together the Heterogeneous Party: A Complementary Social Data Science Experiment." *Big Data & Society* 4 (2):1–15. Accessed April 17, 2021. https://doi.org/10.1177/2053951717736337.

Blok, Anders, and Torben Elgaard Jensen. 2011a. *Bruno Latour: Hybrid Thoughts in a Hybrid World*. London: Routledge.

Blok, Anders, and Torben Elgaard Jensen. 2011b. "We Would like to Do a Bit of Science Studies with You… An Interview with Bruno Latour." In *Bruno Latour: Hybrid Thoughts in a Hybrid World*, edited by Andres Blok, and Torben Elgaard Jensen, 151–66. London: Routledge.

Bornakke, Tobias, and Brian L Due. 2018. "Big–Thick Blending: A Method for Mixing Analytical Insights from Big and Thick Data Sources." *Big Data & Society* 5 (1): 1–16. Accessed April 17, 2021. https://doi.org/10.1177/2053951718765026.

Brown, Phil. 1997. "Popular Epidemiology Revisited." *Current Sociology* 45 (3): 137–56. Accessed April 17, 2021. https://doi.org/10.1177/001139297045003008.

Calvillo, Nerea. 2018. "Particular Sensibilities." *E-Flux Architecture*. Accessed April 17, 2021. https://www.e-flux.com/architecture/accumulation/217054/particular-sensibilities/.

Calvillo, Nerea and Garnett, Emma. 2019. "Data Intimacies: Exploring Embodied Encounters with Air Pollution." *The Sociological Review* 67 (2) 340–356.

Charles, Vincent, and Tatiana Gherman. 2019. "Big Data Analytics and Ethnography: Together for the Greater Good." In *Big Data for the Greater Good*, edited by Ali Emrouznejad, and Vincent Charles, 19–33. Accessed May 3, 2021. https://doi.org/10.1007/978-3-319-93061-9_2.

Chiusi, Fabio, Sarah Fischer, Nicolas Kayser-Bril, and Matthias Spielkamp. 2020. "Automating Society Report 2020." https://automatingsociety.algorithmwatch.org.

Coopmans, Catalijne, Janet Vertesi, Michael Lynch, and Steve Woolgar, eds. 2014. *Representation in Scientific Practice Revisited*. Cambridge, MA: MIT Press.

Crawford, Kate, and Vladan Joler. 2018. "Anatomy of an AI System." *Anatomy of an AI System*. Accessed April 27, 2021. http://www.anatomyof.ai.

CRC Media of Cooperation. 2020. "Research Program." *CRC Media of Cooperation*. Accessed April 27, 202 1. https://www.mediacoop.uni-siegen.de/en/forschungsprogramm/.

Daston, Lorraine, and Peter Gallison. 2007. *Objectivity*. Brooklyn, NY: Zone Books.

Denzin, Norman K., and Yvonna S. Lincoln. 2005. "Introduction: The Discipline and Practice of Qualitative Research." In *The SAGE Handbook of Qualitative Research*, edited by Denzin Norman K., and Yvonna S. Lincoln, 1–32. Thousand Oaks, CA: SAGE Publications.

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

DiSalvo, Carl. 2015. *Adversarial Design*. Cambridge, MA: The MIT Press.

Dumit, Joseph. 2004. *Picturing Personhood: Brain Scans and Biomedical Identity*. Princeton, NJ: Princeton University Press.

Estalella, Adolfo, and Tomas Sanchez Criado, eds. 2018. *Experimental Collaborations: Ethnography through Fieldwork Devices*. New York: Berghahn Books.

Felten, Ed. 2014. "Facebook's Emotional Manipulation Study: When Ethical Worlds Collide." *Freedom to Tinker*. Accessed May 3, 2021. https://freedom-to-tinker.com/2014/06/30/facebooks-emotional-manipulation-study-when-ethical-worlds-collide/.

Flick, Catherine. 2016. "Informed Consent and the Facebook Emotional Manipulation Study." *Research Ethics* 12 (1): 14–28. Accessed May 3, 2021. https://doi.org/10.1177/1747016115599568.

Fortun, Kim. 2001. *Advocacy after Bhopal: Environmentalism, Disaster, New Global Orders*. Chicago: University of Chicago Press. Accessed May 3, 2021. https://press.uchicago.edu/ucp/books/book/chicago/A/bo3641096.html.

Fortun, Kim. 2012. "Ethnography in Late Industrialism." *Cultural Anthropology* 27 (3): 446–64. Accessed May 3, 2021. https://doi.org/10.1111/j.1548-1360.2012.01153.x.

Fortun, Mike, and Kim Fortun. 2013. "The Asthma Files." *The Asthma Files*. Accessed May 3, 2021. https://theasthmafiles.org/.

Fortun, Mike, Kim Fortun, and George E. Marcus. 2016. "Computers in/and Anthropology: The Poetics and Politics of Digitization." In *The Routledge Companion to Digital Ethnography*, edited by Larissa Hjorth, Heather A. Horst, Anne Galloway, and Bell Genevieve, 11–21. New York/London: Routledge. Accessed May 3, 2021. https://doi.org/10.4324/9781315673974.

Fourcade, Marion, and Kieran Healy. 2017. "Seeing like a Market." *Socio-Economic Review* 15 (1): 9–29. Accessed May 3, 2021. https://doi.org/10.1093/ser/mww033.

Gabrys, Jennifer, Helen Pritchard, and Benjamin Barratt. 2016. "Just Good Enough Data: Figuring Data Citizenships through Air Pollution Sensing and Data Stories." *Big Data & Society* 3 (2): 1–14. Accessed May 3, 2021. https://doi.org/10.1177/2053951716679677.

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

Garnett, Emma. 2016. "Developing a Feeling for Error: Practices of Monitoring and Modelling Air Pollution Data." *Big Data & Society* 3 (2): 1–12. Accessed May 3, 2021. https://doi.org/10.1177/2053951716658061.

Garnett, Emma. 2017. "Air Pollution in the Making: Multiplicity and Difference in Interdisciplinary Data Practices." *Science, Technology, & Human Values*, (42) 5: 901–924. Accessed May 3, 2021. https://doi.org/10.1177/0162243917699974.

Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12 (3): 347–64. Accessed May 3, 2021. https://doi.org/10.1177/1461444809342738.

Hine, Christine, ed. 2005. *Virtual Methods: Issues in Social Research on the Internet*. Oxford, New York: Berg.

Hjorth, Larissa, Heather A. Horst, Anne Galloway, and Bell Genevieve, eds. 2016. *The Routledge Companion to Digital Ethnography*. Routledge Handbooks Online. Accessed May 4, 2021. https://doi.org/10.4324/9781315673974.

Holmes, Douglas R., and George E. Marcus. 2008. "Para-Ethnography." In *The SAGE Encyclopedia of Qualitative Research Methods*, edited by Lisa Given, 596–97. Thousand Oaks: SAGE Publications. Accessed May 4, 2021. https://doi.org/10.4135/9781412963909.n307.

Laaksonen, Salla-Maaria, Matti Nelimarkka, Mari Tuokko, Mari Marttila, Arto Kekkonen, and Mikko Villi. 2017. "Working the Fields of Big Data: Using Big-Data-Augmented Online Ethnography to Study Candidate–Candidate Interaction at Election Time." *Journal of Information Technology & Politics* 14 (2): 110–31. Accessed May 5, 2021. https://doi.org/10.1080/19331681.2016.1266981.

Latour, Bruno, and Emilie Hermant. 2006. "Paris: Invisible City." *Bruno-Latour.fr*. Accessed May 5, 2021. http://www.bruno-latour.fr/virtual/EN/index.html.

Law, John. 2004. *After Method: Mess in Social Science Research*. International Library of Sociology. London/New York: Routledge.

Lehtiniemi, Tuuka, and Minna Ruckenstein. 2022. "Prisoners Training AI: Ghosts, Humans and Values in Data Labour." In *Everyday Automation: Experiencing and Anticipating Emerging Technologies*, edited by Sarah Pink, Martin Berg, Deborah Lupton, and Minna Ruckenstein, 184–196. London: Routledge.

Lindgren, Simon. 2020. *Data Theory: Interpretive Sociology and Computational Methods*. Cambridge/Medford, MA: Polity.

Lury, Celia, and Nina Wakeford, eds. 2012. *Inventive Methods: The Happening of the Social*. London/New York: Routledge.

Marcus, George E. 1995. "Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography." *Annual Review of Anthropology* 24 (1): 95–117.

Marcus, George E. 2011. "Multi-Sited Ethnography: Five or Six Things I Know About It Now." In *Multi-Sited Ethnography: Problems and Possibilities in the Translocation of Research Methods*, edited by Simon Coleman and Pauline von Hellermann, 16–34. New York/London: Routledge. Accessed May 5, 2021. https://doi.org/10.4324/9780203810156.

Markham, Annette N. 2005. "The Methods, Politics, and Ethics of Representation in Online Ethnography." In *The Sage Handbook of Qualitative Research*, edited by Denzin Norman K., and Yvonna S. Lincoln, 793–820. Thousand Oaks, CA: SAGE Publications.

Marres, Noortje. 2017. *Digital Sociology: The Reinvention of Social Research*. Cambridge/Malden, MA: Polity.

Marres, Noortje, and Carolin Gerlitz. 2016. "Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology." *The Sociological Review* 64 (1): 21–46. Accessed May 5, 2021. https://doi.org/10.1111/1467-954X.12314.

Marres, Noortje, and Esther Weltevrede. 2013. "Scraping the Social?" *Journal of Cultural Economy* 6 (3): 313–35. Accessed May 5, 2021. https://doi.org/10.1080/17530350.2013.772070.

Mauss, Marcel. 2002. *The Gift: The Form and Reason for Exchange in Archaic Societies*. London: Routledge.

Murphy, Michelle. 2006. *Sick Building Syndrome and the Problem of Uncertainty: Environmental Politics, Technoscience, and Women Workers*. Durham, N.C.: Duke University Press.

Myers, Natasha. 2015. *Rendering Life Molecular: Models, Modelers, and Excitable Matter. Experimental Futures: Technological Lives, Scientific Arts, Anthropological Voices.* Durham: Duke University Press.

Nafus, Dawn. 2018. "Working Ethnographically with Sensor Data." In *Ethnography for a Data-Saturated World*, edited by Hannah Knox and Dawn Nafus, 233–52.

Manchester: Manchester University Press. Accessed May 5, 2021. https://doi.org/10.7765/9781526127600.00019.

Nafus, Dawn 2019. "Data Aggregation as Social Relations: Making Datasets from Self-Tracking Data." *European Review* 27 (3): 440–54. Accessed May 5, 2021. https://doi.org/10.1017/S106279871900005X.

Nafus, Dawn, and Hannah Knox. 2018. *Ethnography for a Data-Saturated World*. Baltimore, Maryland: Manchester University Press. Accessed May 5, 2021. https://muse.jhu.edu/book/67317/.

Neyland, Daniel. 2008. *Organizational Ethnography*. London: SAGE Publications.

Pantzar, Mika, and Minna Ruckenstein. 2017. "Living the Metrics: Self-Tracking and Situated Objectivity." *Digital Health* 3 (1):1–10. Accessed May 5, 2021. https://doi.org/10.1177/2055207617712590.

Pantzar, Mika, Minna Ruckenstein, and Veera Mustonen. 2017. "Social Rhythms of the Heart." *Health Sociology Review* 26 (1): 22–37. Accessed May 5, 2021. https://doi.org/10.1080/14461242.2016.1184580.

Paßmann, Johannes, and Cornelius Schubert. 2020. "Technografie als Methode der Social-Media-Forschung." In *Handbuch Diskurse Digita*l. Berlin: De Gruyter. Preprint, submitted April 20, 2020. Accessed May 5, 2021. https://doi.org/10.13140/RG.2.2.22738.50883.

Pink, Sarah, Heather A. Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi, eds. 2016. *Digital Ethnography: Principles and Practice*. Los Angeles: SAGE.

Pink, Sarah, Deborah Lupton, Martin Berg, Paul Dourish, Adrian Dyer, Vaike Fors, Edgar Gómez Cruz, et al. 2016. "Data Ethnographies (1): Personal Data in an Uncertain World." Working Paper, submitted January 1, 2016. Accessed May 5, 2021. https://doi.org/10.13140/RG.2.2.27713.76643.

Pink, Sarah, Minna Ruckenstein, Robert Willim, and Melisa Duque. 2018. "Broken Data: Conceptualising Data in an Emerging World." *Big Data & Society* 5 (1): 1–13. Accessed May 5, 2021. https://doi.org/10.1177/2053951717753228.

Ruckenstein, Minna. 2014. "Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles." *Societies* 4 (1): 68–84. Accessed May 5, 2021. https://doi.org/10.3390/soc4010068.

Ruckenstein, Minna. 2019. "Tracing Medicinal Agencies: Antidepressants and Life-Effects." *Social Science & Medicine* 235 (August): 1–9. Accessed May 5, 2021. https://doi.org/10.1016/j.socscimed.2019.112368.

Ruckenstein, Minna, and Julia Granroth. 2020. "Algorithms, Advertising and the Intimacy of Surveillance." *Journal of Cultural Economy* 13 (1): 12–24. Accessed May 5, 2021. https://doi.org/10.1080/17530350.2019.1574866.

Ruckenstein, Minna, and Linda Lisa Maria Turunen. 2020. "Re-Humanizing the Platform: Content Moderators and the Logic of Care." *New Media & Society* 22 (6): 1026–42. Accessed May 5, 2021. https://doi.org/10.1177/1461444819875990.

Ruckenstein, Minna, and Trifuljesko, Sonja. Forthcoming. "Algorithmic futures and the unsettled sense of care." In *An Anthropology of Futures and Technologies*, edited by Debora Lanzeni, Sarah Pink, Karen Waltorp, and Rachel Charlotte Smith. London/New York: Routledge.

Ruppert, Evelyn, John Law, and Mike Savage. 2013. "Reassembling Social Science Methods: The Challenge of Digital Devices." *Theory, Culture & Society* 30 (4): 22–46. Accessed May 5, 2021. https://doi.org/10.1177/0263276413484941.

Schwartz, Howard. 2002. "Data: Who Needs It? Describing Normal Environments—Examples and Methods." *Ethnographic Studies* 2002 (7): 7–32.

Star, Susan Leigh. 1999. "The Ethnography of Infrastructure." *American Behavioral Scientist* 43 (3): 377–91. Accessed May 5, 2021. https://doi.org/10.1177/00027649921955326.

Thompson, Craig J. 2019. "The 'Big Data' Myth and the Pitfalls of 'Thick Data' Opportunism: On the Need for a Different Ontology of Markets and Consumption." *Journal of Marketing Management* 35 (3–4): 207–30. Accessed May 5, 2021. https://doi.org/10.1080/0267257X.2019.1579751.

Tsing, Anna Lowenhaupt. 2015. *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Princeton, NJ: Princeton University Press.

Venturini, Tommaso, Liliana Bounegru, Jonathan Gray, and Richard Rogers. 2018. "A Reality Check (List) for Digital Methods." *New Media & Society* 20 (11): 1–23. Accessed May 5, 2021. https://doi.org/10.1177/1461444818769236.

Venturini, Tommaso, Mathieu Jacomy, Axel Meunier, and Bruno Latour. 2017. "An Unexpected Journey: A Few Lessons from Sciences Po Médialab's Experience." *Big Data & Society* 4 (2): 1–11. Accessed May 5, 2021. https://doi.org/10.1177/2053951717720949.

Venturini, Tommaso, and Richard Rogers. 2019. "'API-Based Research' or How Can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach." *Digital Journalism* 7 (4): 532–40. Accessed May 5, 2021. https://doi.org/10.1080/21670811.2019.1591927.

Watts, Laura, and Dawn Nafus. 2013. *Data Stories*. Orkney: Brae Editions.

Woolgar, Steve. 1990. "Configuring the User: The Case of Usability Trials." *The Sociological Review* 38 (1_suppl): 58–99. Accessed May 5, 2021. https://doi.org/10.1111%2Fj.1467-954X.1990.tb03349.x.

Woolgar Steve, Noortje Marres, and Javier Lezaun, et al. 2009. "From scale to scalography: a provocation piece." *International workshop on Scalography*, Oxford, 8 July 2009, p. 7. Saïd Business School, University of Oxford.

Ziewitz, Malte. 2017a. "Experience in Action: Moderating Care in Web-Based Patient Feedback." *Social Science & Medicine* 175 (February): 99–108. Accessed May 5, 2021. https://doi.org/10.1016/j.socscimed.2016.12.028.

Ziewitz, Malte. 2017b. "A Not Quite Random Walk: Experimenting with the Ethnomethods of the Algorithm." *Big Data & Society* 4 (2): 1–13. Accessed May 5, 2021. https://doi.org/10.1177/2053951717738105.

Ziewitz, Malte. 2019. "Rethinking Gaming: The Ethical Work of Optimization in Web Search Engines." *Social Studies of Science* 49 (5): 707–31. Accessed May 5, 2021. https://doi.org/10.1177/0306312719865607.

Ziewitz, Malte, and Ranjit Singh. 2021. "Critical Companionship: Some Sensibilities for Studying the Lived Experience of Data Subjects." *Big Data & Society* 8 (2): 20539517211061120. Accessed April 2, 2022. https://doi.org/10.1177/20539517211061122.