Marcus Burkhardt, Daniela van Geenen,
Carolin Gerlitz, Sam Hind, Timo Kaerlein,
Danny Lämmerhirt, Axel Volmar (eds.)

INTERROGATING DATAFICATION

Towards a Praxeology of Data

[transcript] Media in Action

Marcus Burkhardt, Daniela van Geenen, Carolin Gerlitz, Sam Hind, Timo Kaerlein, Danny Lämmerhirt, Axel Volmar (eds.)
Interrogating Datafication

Media in Action | Volume 3

Marcus Burkhardt is a lecturer in the media studies department at the University of Siegen. Together with Karin Knorr Cetina, he heads the project "Agentic Media: Formations of Semi-Autonomy" in the DFG Collaborative Research Centre 1187 "Media of Cooperation". He has a background in media studies, philosophy, and computer science and received his doctorate at Justus-Liebig-University Gießen with a thesis on the media theory of digital databases. His research interests include the history and theory of digital media, especially the logi(sti)cs of database technologies, big data, and algorithmic media as well as digital methods.

Daniela van Geenen is a Ph.D. candidate at the DFG Collaborative Research Center 1187 "Media of Cooperation" at the University of Siegen, after receiving her MA from Utrecht University. Her research lies at the interface of media studies and science and technology studies (STS), with particular focus on (critical) data studies, software, device, and infrastructure studies. In her Ph.D. project she investigates the sociomaterial organization, (data) practices, and values in/of urban sensing, paying specific attention to cases of environmental sensing in relation to mobility monitoring and planning. She is also a lecturer in Data Journalism and Visualization at the University of Applied Sciences Utrecht.

Carolin Gerlitz is a professor of digital media and methods at the University of Siegen and member of the Digital Methods Initiative Amsterdam. She is co-speaker of the DFG Collaborative Research Centre 1187 "Media of Cooperation" and the DFG Graduate School "Locating Media". Her research interests are social media and platform studies, app studies, inventive digital methods, sensor media, quantification, automation, and issue mapping.

Sam Hind is a research associate in the DFG Collaborative Research Centre 1187 "Media of Cooperation" at the University of Siegen. His research currently revolves around sensor-based navigation as part of the "Navigation in Online/Offline Spaces" project. He received his PhD in the Centre for Interdisciplinary Methodologies at the University of Warwick, UK. His research interests include digital navigation, mapping practices and mobility.

Timo Kaerlein is a lecturer at the Institute for Media Studies at the Ruhr University Bochum. He has a background in media studies and received his doctorate at Paderborn University in 2017. His research focuses on the theory, history, and aesthetics of interfaces, mobile media, social robotics, and media cultures of obsolescence.

Danny Lämmerhirt is a PhD-Candidate at the Graduate School (University of Siegen) with an interest in the politics of participatory data infrastructures, data governance, quantification, and valuation. His dissertation compares different citizen-centered data sharing models for digital health data, exploring how these models enable cooperation, governance, and the valuation of health data along multiple (possibly conflicting) interests.

Axel Volmar is a postdoctoral researcher at the Collaborative Research Center "Media of Cooperation" at the University of Siegen. He published on the history and theory of digital media, sound studies, format theory, and digital temporality.

Marcus Burkhardt, Daniela van Geenen, Carolin Gerlitz, Sam Hind,
Timo Kaerlein, Danny Lämmerhirt, Axel Volmar (eds.)

Interrogating Datafication

Towards a Praxeology of Data

[transcript]

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

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de

First published in 2022 by transcript Verlag, Bielefeld

Cover layout: Maria Arndt, Bielefeld
Cover illustration: www.vectorstock.com
Printed by Majuskel Medienproduktion GmbH, Wetzlar
Print-ISBN 978-3-8376-5561-2
PDF-ISBN 978-3-8394-5561-6
https://doi.org/10.14361/9783839455616
ISSN of series: 2749-9960
eISSN of series: 2749-9979

Printed on permanent acid-free text paper.

Contents

III: Digital Care

IV: Datafied Mobilities

Epilogue

Introduction

Marcus Burkhardt, Daniela van Geenen, Carolin Gerlitz, Sam Hind, Timo Kaerlein, Danny Lämmerhirt, and Axel Volmar

The editing process of the present volume took place in the midst of the global COVID-19 pandemic. This historical moment has been accompanied by a marked shift in public perceptions of the role data plays in politics and everyday life. We have witnessed the emergence of highly visible data-intensive infrastructures and control mechanisms that have been developed to keep track of the spread of the pandemic, mitigating its effects locally and globally, and modulating established behavioral routines. Governments around the world turned to experimenting with a range of digital tracking tools and automated decision-making systems (Algorithm Watch 2020), giving rise to a new form of "sensory power" enacted through sensory assemblages targeting infectious "clusters as objects of government," including "hotspots, epicentres and bubbles" (Isin and Ruppert 2020, 8). Private sector parties were keen to offer their expertise and existing infrastructures and especially big tech platforms such as Apple and Google acted as "gatekeepers," for instance, governing and restricting app developers' access to Apple's App Store and Google's Playstore (Dieter et al. 2021). In doing so, the rise of COVID-19 pandemic response apps exacerbated public sector organizations' dependencies on private sector infrastructures. In short, the felt urgency to develop technological solutions brought on by the spread of a deadly virus has acted to catalyze and amplify ongoing developments of "datafication" (cf. Mayer-Schönberger and Cukier 2013; van Dijck 2014) that have been profoundly shaping society since at least the 1980s.

Mayer-Schönberger and Cukier first coined the term "datafication" to signal a general "transformation in how society processes information" (Mayer-Schönberger and Cukier 2013, 29), in which ever increasing areas of social, political, and economic life, including everyday administrative processes, healthcare, education, daily news coverage, cultural and industrial production were transformed by the increasingly automated generation and processing of data. In this urge to render all manner of activities into machine-readable formats – say, reading the news, listening to music, or hiring a bike – and producing continuous Big Data flows, activities are not simply granted a representational shadow or "data double" (Ruckenstein 2014) but are fundamentally re-constituted in this process. In becoming "datafied" versions of their former selves, these activities transform.

It would be short-sighted, however, to situate the pandemic intensification of practices of datafication exclusively on the side of (state) governance and "platformized" forms of control (cf. Poell et al. 2019), acting as top-down mechanisms to monitor, regulate and/or monetize the activities and behaviors of populations. On the contrary, the pandemic has equally foregrounded *everyday engagements with data* (cf. Pink et al. 2017 on "mundane data," Smith 2018 on "data doxa"), beyond state-led statistical interventions (Ruppert and Scheel 2021). The German Corona-Datenspende App, to focus on one striking example, encapsulates various aspects of society's shifting practices with data. Not only is it an example of a new kind of "data altruism" (cf. European Commission 2020) in which the app reifies personal data as an object to be given away for (speculative) public benefit. It also perpetuates longstanding promises of prediction that the combination of novel data sources such as wearable sensor-based, algorithmic technologies and devices enable.

Likewise, the pandemic has challenged our capabilities to make sense of data (Leonelli 2021). Corona dashboards combine various kinds of (possibly competing) knowledges as well as methods for data acquisition, analysis, and representation, frequently driving analysts and decision-makers down into "rabbit holes of investigating different metrics and standards" (Correll and Froehlich 2020, n.p.). Numbers, statistical values, and indicators proliferate together with multiple ways of counting, monitoring, and enacting the pandemic (cf. Day, Lury, and Wakeford 2014). The challenges and choices of enumerating the pandemic – also addressed and problematized as an "infodemic" (World Health Organization 2020), in which the authority of knowledge produced by public institutions is put into question – do not only affect lay people. Competing ways of counting and acting upon COVID-19 data have also foregrounded professional and expert disagreements and uncertainties (Ruppert, Isin, and Bigo 2019) as a critical challenge for the management of the pandemic.

Incidence rates in particular, often expressed as the number of cases per 100,000 over a seven-day period, have come to be seen as a key metric for worldwide monitoring and comparison, for instance, in European politics and media coverage (e.g., in The Economist's [2021] section "Tracking the coronavirus across Europe"). The result, or at least an *intended promise*, of such a metric has been to offer a standardized mechanism through which people are able to respond accordingly: adapting their behavior in light of the numbers. A popular slogan also emerged in the early stages of the pandemic: *Flatten the Curve!* instructing people to act responsibly to help reduce incidence rates.[1] Imposing a normative shortcut between data representations and everyday practices, citizens were expected to consider appropriate levels of personal hygiene. Moreover, social events were conducted on the basis of the incidence curve. Yet, the attendant promise – of citizens responding effectively and diligently to rising incidence rates by modifying their behavior, decisions, and movements – has not necessarily been matched by reality. Indeed, as the pandemic continued to rage, people's capacity – and often, willingness – to flatten the curve waned, dependent not only on changing circumstances (e.g., the closing of kindergartens or the re-opening of workplaces) but also on the fluctuating relevance of specific metrics from incidence rates to *r* values to hospital bed capacities.[2] What we have seen emerge, then, is a far messier array of *data practices* linked to these metrics; entangled in the pandemic-specific social lives of citizens.

The condensation of the pandemic moment has afforded ample opportunity to re-examine and interrogate how processes of datafication operate. Accordingly, the goal of the present volume is twofold: 1) to understand processes of datafication as grounded in and composed of heterogeneous practices of data creation, collection, cleaning, processing, analysis, archiving, transfer and re-use, among others, and 2) to scrutinize how processes of datafication increasingly target fluid, mobile and ephemeral phenomena, e.g., in the capturing of local and real-time transactional data generated through everyday practices (cf. Agre 1994). By stressing the role of situated practices within and throughout macro processes of datafication, we follow the premise that the social is always enacted in "practical accomplishments" (Garfinkel 1967, 9) and ongoing acts of mutual sense-making. To put it succinctly, datafication does not just happen on its own, but is manifested through everyday interactions between people, infrastructures, and established conventions.

The present volume argues that in order to understand how datafication continues to redefine societies epistemologically, economically, and socially, we need to turn our scholarly attention towards *practices*. As a macro-phenomenon datafication has the potential to appear abstract, despite its obvious entanglement with local contexts of use and specific *communities of practice* (Lave and Wenger 1991). In this, practices of producing, collecting, aggregating, disseminating, processing,

1 See, for instance, this news coverage in *The New Yorker* from 27 March 2020, the early stages of the COVID-19 pandemic: https://www.nytimes.com/article/flatten-curve-coronavirus.html.

2 This is, for example, indicated by mobility measures such as monitored by the COVID-19 Mobility Project based on anonymized mobile phone data: https://www.covid-19-mobility.org/mobility-monitor/. For an impression of shifting attitudes and behavior toward official COVID-19 measures see, for instance, the German COVID-19 Snapshot Monitoring (COSMO): https://projekte.uni-erfurt.de/cosmo2020/web/.

representing and displaying, analyzing and re-using data become vital in exploring and accounting for how datafication operates. We are thus interested in the *practical accomplishment of datafication*, including hard-to-observe data work taking place behind digital media. As Clare Southerton (2020, 3) writes, "[w]ith the mass infiltration of smart technologies into everyday life and as more social interaction is filtered through social media platforms and other online services, data is now generated and collected from a diverse array of practices." Thus, as the volume explores, datafication is a pervasive phenomenon, occurring far beyond social media platforms within the extended "digital enclosures" (Andrejevic 2007, 297) of smart cities and homes, and always-on mobile devices. In today's digital media environments, it is hard to find a practice that is not datafied to some extent. But while data are increasingly produced in all areas of social life, datafication is by no means just an automatic process. On the contrary, it has to be considered as the result of practical work.

In the following, we will briefly summarize the current state of research around issues of datafication in the humanities and social sciences. In particular, we highlight various contributions in the fields of (critical) data studies and offer historical perspectives on (digital) data to develop a more nuanced understanding of datafication, focusing on the situated character of *all* data. That is, data is produced under specific local conditions, processed and (re)appropriated in heterogeneous situations of use. Building on this existing work, we develop our own proposal for a praxeology of data in the next part of the introduction by drawing on the rich tradition of research inspired by ethnomethodology and emerging ethnographic accounts of data practices. The introduction concludes with an outline of the structure of the volume itself.

Datafication: Operations, Logics, and Critiques

Rephrasing Mayer-Schönberger and Cukier's definition of datafication, José van Dijck notes that datafication is the "transformation of social action into online quantified data" predicated on "real-time tracking and predictive analysis" (van Dijck 2014, 198). In this, datafication is cast as a "legitimate means to *access*, *understand*, and *monitor* people's behaviour" (2014, 198, author's emphasis) online. Van Dijck, writing in 2014, discusses datafication in the context of social media, focusing on a roster of platforms, now commonplace, from Facebook to YouTube. Real-time tracking of user activity on these "data-intensive" platforms (Gerlitz and Helmond 2013, 1349), ensures all manner of comments, likes, tags, uploads, edits, and other similar interactions are recorded. Following the logic of the "Like economy" (2013, 1349), datafication provides an opportunity to capture, and increasingly extract, *forms of value* from users.

In the context of social media and communication platforms, datafication is presumed to be a silent, discrete, unobtrusive process in which users of such services are largely unaware of how their interactions are being datafied. We argue, however, that datafication implies two relations between data and practices: both accomplished *through* practices while also *capturing* specific practices. In the context of social media, Ganaele Langlois et al. (2009) have argued that social media usage such as posting or liking content involves both social as well as data practices. At the very least, platforms may facilitate user (inter)actions, with users free to use these services (e.g., reading and commenting on (blog) posts, sending messages, listening to and recommending music etc.) unaware of background, or "back-end" (Gerlitz 2016, 28) datafication processes. Yet, as we hope to show throughout this collection, this is far from the case in the contemporary digital media landscape and the platformized technological infrastructures it thrives on, in which datafication processes are the basis for interactions and engagements of diverse user groups with recreational or professional interests. Included in the latter group are third party "complementors" who offer their services making use of, and are therefore dependent on, platform data infrastructures and modes of governance (Poell et al. 2019, 6–9). Such datafied modes of platform governance may involve continuous data capture in the background or passively through the integration of automated sensing processes (cf. Thielmann 2019). Rather often, though, platform governance includes mechanisms and modalities of surfacing data from and for diverse users in the form of "participative metrics" (Gerlitz and Lury 2014, 174) such as scores, rankings, and ratings, which in turn both exploit and shape the practices of these users (Fourcade and Healy 2017).

In organizational and technical terms, then, datafication is associated with "platformization," which has been "defined as the penetration of infrastructures, economic processes and governmental frameworks of digital platforms in different economic sectors and spheres of life" (Poell et al. 2019, 1). Drawing from software studies, business studies, critical political economy, and cultural studies, Poell et al. (2019, 3) "define platforms as (re-)programmable digital infrastructures that facilitate and shape personalised interactions among end-users and complementors, organised through the systematic collection, algorithmic processing, monetisation, and circulation of data." The influence of the "big five" platforms – Google, Apple, Facebook, Amazon, and Microsoft – have marked today's "platform society" (van Dijck, Poell, and de Waal 2018) in which a few platform businesses offer all kinds of services from online teaching to housing. As a result, Poell et al. (2019, 5, our emphasis) "stress the importance of considering *platform-based user practices*" striving to "trace how institutional changes and *shifting cultural practices* mutually articulate each other." In this, data generated through platform-based user practices simply serves to further embed platforms in everyday life.

Van Dijck's (2014) early critique of datafication can be seen as a kind of pre-empirical intervention, laying out the foundational principles of its operation (i.e. accessing, understanding, and monitoring user actions and behavior) as it was then found across social media platforms. In the years since, however, datafication has arguably become more *forceful*, being imposed in and on other realms previously untouched by it, and offering a means to *intervene*. In an economic environment revolving around the monetization of data, processes of datafication have become increasingly *unavoidable* not only for certain kinds of online users, but a whole host of people from school children (e.g., Kerssens and van Dijck 2021; Williamson 2016) to delivery bike couriers and drivers (e.g., Shapiro 2020; Pentenrieder, this volume) and pedestrians (e.g., Mattern 2014, O'Grady, this volume). Naturally, there are also geographical, political, and legal differences in the *spread* and *depth* of datafication globally, see for instance the effect of the EU's GDPR (General Data Protection Regulation) on data protection and privacy. Whilst the premises of datafication may indeed be shared – even between the USA, Europe, and China – on the ground realities differ markedly and are deserving of systematic analyses.

To reiterate, it is becoming clear that datafication is a plural phenomenon, dependent on *who* is doing the datafication. Besides specific platforms, the big five are variously involved in gleaning understanding of and regulating user behavior, using different strategies and techniques to do so. Away from big tech, many other corporate parties are striving for datafication, such as those in the healthcare sector (e.g., 23andMe), the agriculture sector (Bayer Crop Science), or the automotive industry (e.g., Hind, this volume). In these application contexts, datafication is bound to look differently, adapting to particular industry demands and expectations, as well as navigating different legal restrictions and cultural discourses. The question, then, is the extent to which these *datafications* do indeed share common aspects and logics. Yet, as van Dijck has already underscored, "all three apparatuses – corporate, academic, and state – are highly staked in getting unrestrained access to metadata as well as in the public's acceptance of datafication as a leading paradigm" (van Dijck 2014, 203). In this, datafication is not simply a technical procedure (of collecting and processing data), a social phenomenon (attempting to capture and shaping behavior), or an economic process (generating profit) but a *political project* motivating the whole gamut of social actors from global corporations to nation-states.

Also returning to Mayer-Schönberger and Cukier, Ulises A. Mejias and Nick Couldry contend that datafication is not merely digitization because, through the former "large domains of human life bec[ome] susceptible to being processed via forms of analysis that [can] be automated on a large-scale" (Mejias and Couldry 2019, 2). In this, they offer a similar definition of datafication to van Dijck: "the transformation of human life into data through quantification" (2019, 3), emphasizing not how datafication becomes a means through which to merely access or understand people's behavior, but how it becomes a means to actively *shape* and *manage* social activities. In this, Meijas and Couldry suggest that datafication is enabled through particular *infrastructures* such that "life actions previously performed elsewhere (such as communicating with friends, sharing cultural products, hailing a taxi etc.)…" (2019, 3) become *re-routed* and *re-organized*.

Nonetheless, writing with a far greater sense of urgency, Couldry also suggests that there is a need to "grasp a world where a general project of social reconstruction…is under way" (Couldry 2020, 1140). Whilst Mejias and Couldry write that datafication involves the transformation of human life, they also add that it entails "the generation of different kinds of value from data" (Meijas and Couldry 2020, 3). Yet, between the lines, it is evident that datafication is not necessarily generating different kinds of value at all, but quite a narrow, quantifiable kind of value derived from captured data, and captured data only. As they later posit: "[e]ven more importantly" than the large-scale re-organization of social actions and activities, "the process of quantification involves *abstraction* via the process of turning the flow of social life and social meaning into streams of numbers that can be counted" (2020, 3, author's emphasis). That which cannot be counted, following the logic of datafication, does not count. Put otherwise, the question is "[w]hat counts and who counts" (Gerlitz 2016, 33). Aspects of social life that cannot, or *will not*, be transformed may escape datafication, but in a world being rapidly datafied the risk of *not* being transformed may perhaps be greater; dismissed as ill-fitting, unruly, or irrelevant.

With this realization, Couldry wonders whether some philosophical traditions are well-equipped enough to make sense of the "new and radical forms of reduction" (Couldry 2020, 1140–1141) offered by datafication. In particular, he asks whether the "descriptivism" (2020, 1140) of certain approaches such as actor-network theory (ANT), "lose[s] sight of the critical question" (2020, 1140) of what happens when social life either becomes datafied, or dismissed. In a similar fashion, Mirko Tobias Schäfer and Karin van Es (2017) ask whether studying culture through data is possible and desirable, and what kinds of epistemological reflection and methodological criticism such approaches offer. The following paragraphs offer different strategies to keep the critical question in sight situating data in society, including from historical perspectives.

Situating Data in Society

Critical Data Studies (CDS) emerged as a reaction to the propagation of processes of datafication and their product Big Data, both as sociotechnical phenomena and epistemological promises (e.g., boyd and Crawford 2012; Kitchin 2014a; 2014b; Iliadis and Russo 2016). danah boyd and Kate Crawford's (2012) article can be seen as seminal, posing “critical questions for Big Data,” taken up in subsequent publications (cf. Dalton, Taylor, and Thatcher 2016; Iliadis and Russo 2016; Kitchin and Lauriault 2018). In it, boyd and Crawford (2012) discuss the required (preliminary) knowledge and expertise for understanding Big Data, the (in)accessibility – and (in)assessability – of data, including involved methodological assumptions, technical infrastructures, and tools. Their aim was to evaluate and reflect on actual possibilities in comparison to the “mythology” (2012, 662) of Big Data, or the positivist empiricism propagated by those who celebrated the “accuracy” and “objectivity” of large-scale, automated data collection.

In an influential dialog, Craig Dalton, Linnet Taylor and Jim Thatcher suggested how scholars could develop and express “critical agendas and responses” to “data and algorithmic analytics” (2016, 1). In this, the emergent interdisciplinary field of CDS could serve to draw together “diverse sets of work around data’s recursive relationship to society” (2016, 1), driven by a collective interest in the broader social embeddedness of digital data, and the techniques involved in their production and distribution. In short, that data has significant cultural, historical, economic, and political qualities worthy of specific focus. Yet, Dalton, Taylor, and Thatcher also addressed the challenge of CDS to connect those who “use critical theory [to] those who engage in rigorous empirical research” (2016, 1), advocating the need to establish a dialog between conceptual and applied work. Craig Dalton and Jim Thatcher’s original call for CDS also understood data as inherently spatial, requiring understanding the “contextual value” of place (2014, n.p.): that data need to be considered *in situ*, connecting to the intrinsically socio-technical settings and situations of their making.

In recent years, CDS work has received more criticism, especially on the notion of critique itself. In particular, scholars have criticized the distant, macro-sociological focus of early CDS work mainly tackling the technological infrastructures involved in datafication processes, and less concerned with actual practices of people confronted by and interacting with, data and algorithmic systems (Christin 2017; Dencik 2019; Leonelli 2021). More recent “data studies” approaches have foregrounded empirical and interventionist work. These range from calls to radicalize the scholarly community to become more active and socially engaged, to work that has explicitly investigated the (often silent) voices of practitioners, non-experts, and lay people. For example, Neff et al. (2017) have invited scholars across the data science-CDS divide to work collectively to “help push for more ethical, and better, ways” (2017, 85) to know the “datafied society.” Likewise, Kennedy (2018, 18) has explored the “everyday experiences” of data, and Kennedy et al. (2020, 3) have studied the “public understanding and perceptions” of digital, and personal, data.

These practical critiques of and updates to CDS have explicitly offered activist, feminist, inventive and affective approaches, especially around datafication “from the margins” (Milan, Treré, and Masiero 2021) such as in the Global South (e.g., Crooks and Currie 2021; D'Ignazio and Klein 2020; Marres 2012; Milan and Treré 2019). Recent (critical) data work has also offered novel methodological strategies. For instance, Jo Bates, Yu-Wei Lin, and Paula Goodale (2016) introduced "data journeys" as a methodological device to follow "the life of data" across settings and situations, whilst Nate Tkacz et al. (2021) developed "data diaries" as a similar such device to chart and account for how data "co-constitute a given spatial situation" (2021, 2). Similar such methodological approaches are in evidence in this volume, whether concerning the flow (or not) of data in the context of asylum applications (Al Jaramani, Ponzanesi, and van Schie) or the construction of "data stories" in an academic context (Mosconi et al.).

Historical Perspectives on Data

In a related vein to the contributions in CDS, historians of science and media historians have responded with reservations to accounts that depict datafication as a radical transformation of how data is being processed in societies. Rather than focusing on the alleged "newness" of Big Data, Elena Aronova, Christine von Oertzen, and David Sepkoski (2017) urge data scholars to also consider technological, structural, epistemological, and praxeological continuities that are present in current expressions of datafication. In this respect, they refute claims that present Big Data as a consequence of the digital age by arguing that the "forward-looking rhetoric" of the present discourse tends to conceal that "these technologies have histories, and that those histories stretch back well before the advent of electronic computing" (2017, 2). Against the presentism of common Big Data narratives, they argue for a *longue durée* perspective on datafication since "the project of translating the world into data … has been under way for centuries" (2017, 8). This is evidenced, for instance, in historical practices of data aggregation and database practices, which span back to pre-digital times and material cultures. Although Aronova, von Oertzen, and Sepkoski contend that it is impossible to ignore the impact of electronic computing within the history of data, they strongly warn against "making the introduction of computers a decisive Rubicon in a broader history of data—to avoid, in other words, thinking of data histories as being 'B.C.' (before computers) or 'A.C.'" (2017, 15). In order to stress this understanding this volume also features historical case studies.

Aronova, von Oertzen, and Sepkoski further emphasize that the notion of Big Data is reminiscent of the term Big Science, another capital-letter term that became prominent after World War II to denote the enormous financial, technological and institutional efforts in connection with Cold War science funding: "There are parallels—and indeed direct overlaps—between Big Science and Big Data. Many projects that involve Big Data—the Human Genome Project, CERN, the Very Large

Telescope array—unquestionably fit the definition of Big Science." (Aronova, von Oertzen, and Sepkoski 2017, 3) In both cases, the adjective "big" does not merely refer to the vastness of the data that are being collected and processed but indicates the magnitude of investments being made in both economical and institutional terms.

Against this backdrop, Aronova, von Oertzen, and Sepkoski regard the contemporary phenomenon of Big Data "as a chapter in a longer history (or, rather, histories) of observation, quantification, statistical methods, models, and computing technologies" (2017, 6). At the same time, they stress important differences between past and present forms of datafication as well. While, for instance, the scientific endeavors of data capturing in the pre-electronic era were somewhat "bound in space and time to physical archives and analog infrastructures," the contemporary project of Big Data "radically transcends the circumstances and locality of its production" (Aronova, von Oertzen, and Sepkoski 2017, 16), allowing data sets to move in digital form and to thereby traverse the contexts of their creation at will. The growing mobility and portability of data, in turn, equally raises questions regarding the ownership and provenance of data, and particularly personal data obtained from marginalized populations. More often than not, these questions lead the way once again into forgotten pasts, such as to the common but no less dubious practices of colonial and colonialist data collection (see also de Chadarevian and Porter 2018, 551) and exploitation. For instance, Joanna Radin (2017) tells the story of how a comprehensive long-term dataset on rates of diabetes and obesity in the Native American Akimel O'odham (known in science as Pima) became widely used training data for machine learning applications by means of decontextualization.

Soraya de Chadarevian and Theodore M. Porter (2018, 550) similarly seek to put "into sharper relief the relation of current data practices to earlier ones." Contrary to Aronova, von Oertzen, and Sepkoski, their focus rests less on individuals but foregrounds more "the roles of diverse institutions as sites of data production, including medicine, militaries, industry, commerce, finance, insurance, pensions, libraries, censuses, and bureaus of standards" (2018, 550). Moreover, they consider the central role of technology beyond simply computers, to include "a variety of tools for recording, storing, communicating, and processing information" (2018, 550) and foreground the field of statistics with its double status as both a mathematical and a social field, which "brings out the fundamental and longstanding role of social know-how and state administration in the history of data" (2018, 551).

Nevertheless, de Chadarevian and Porter too acknowledge fundamental differences between scientific data and what they call "social media" data – hereby referring to the digital data "managed by an oligopoly of internet marketing firms that specialize in linking potential customers to taxi rides, hotel stays, and things for sale" (2018, 550). As one of the main differences, they highlight the fact that data in the social web are generally not generated as samples that may serve as a basis for assessing "scientific models or hypotheses" but as personalized indexes used for algorithmic predictions of future behavior: "Such data, in their teeming abundance, have not been content to remain mere samples, but have become universes unto themselves [, supporting] the making of algorithms to anticipate and to nudge future behaviors on the basis of all the numbers generated by previous actions and choices" (2018, 550).

This distinction between the use of data for testing scientific hypotheses and as a means to predict future behaviors of users and customers, however, might also indicate certain limitations of data histories from the history of science. While the existing historiographic scholarship provides productive insights regarding the epistemic dimensions of data practices – including the often problematic relation between quantitative data, the construction of scientific "evidence," and the truth claims that are based on them (de Chadarevian and Porter 2018, 552; see also Leonelli et al. 2017) – the case studies in their totality tend to keep a rather narrow focus on the realm of the sciences and scientific practice. Although many data practices and tools indeed originated in scientific uses, it seems equally important to explore data histories outside the sciences. Karin van Es and Eef Masson (2018), for instance, have approached the history of datafication from the perspective of media studies and media history, with a particular focus on media industries. In a similar vein, this volume features historiographic case studies from the history of cinema and early computing that seek to delineate both the continuities and ruptures of past and contemporary data practices. Taking recent work in data studies as an inspiration, future historiographic research may also study datafication as socio-technical phenomena beyond the domains of both the sciences and the media.

This volume nevertheless follows the path laid out by the more recent science studies with regards to a strong focus on practices. Aronova, von Oertzen, and Sepkoski remind us that data and data practices have often been at the root of controversies over positivist visions of science and scientific progress, as crystallized, most prominently, in Thomas Kuhn's *The Structure of Scientific Revolutions* (1962). In the aftermath of Kuhn's groundbreaking book, science studies "reinvented the history of science as the history of scientific practices rather than scientific ideas" (Aronova, von Oertzen, and Sepkoski 2017, 4). To put it differently: rather than seeing the diverse practices, routines, and tasks that make up laboratory research as negligible procedures that are simply applied to support or falsify theoretical models, post-Kuhnian historians of science and scholars from the emerging fields of science and technology studies (STS) and historical epistemology came to understand them as the very basis of scientific knowledge making and the truth claims often associated with them. These particularly include practices that enable the production and manipulation of data, e.g., through observation and the creation of material traces (see, for instance, Latour 1987). Moreover, practices of data display are vital for rendering data into meaningful forms, such as graphs, diagrams or other forms of visualization (e.g., Rheinberger 2011).

Towards a Praxeology of Data

This volume builds on the socially and historically situated understandings of data outlined above. Yet, as the following sections will show, it takes a distinctive path, offering an account of data *practices*. In this, we take up Lina Dencik's (2019) call to advance a "practice approach to datafication" in order to "consider the *uses* to which data systems are put in social life" (245, author's emphasis). Whilst Dencik herself builds on a Bourdieuian *a priori* notion of practice as "habitus," however, this volume offers a more ethnomethodological (Garfinkel 1967) or "praxeological" (Schüttpelz and Meyer 2017; Gießmann 2018) approach in which data practices are conceived as cooperatively achieved accomplishments.

Noortje Marres (2017), for instance, differentiates three main approaches to studying the social aspects of digital culture: a technology- or platform-centric, a data-centric, and a practice-centric approach. To exemplify, this "social" dimension might respectively be traced back to specific technologies like social media and Web 2.0 (technology/platform), to the capture and processing of data about society (data), or to contexts of action and use (practice). The practice-centered study of digital sociality considers how the use of technology is always *contingent* as people engage with it in myriad situations and settings, from negotiating the Dutch immigration system, to self-monitoring blood pressure. This view also helps to destabilize the ontological security of technologies as singular, fixed entities with specific properties. It draws our attention to how engagements with datafication are not only *practical* accomplishments, but also *distributed* accomplishments (Marres 2012), often involving many connected technologies, and increasingly resulting in "synthetic situations" (Knorr-Cetina 2009).

Our aim is thus to approach data practices as cooperatively performed, articulated and understood through specific and shifting sociomaterial arrangements. The book follows in this spirit of Ruppert and Scheel (2021) who explore the data practices of (international) statisticians, Hobbis (2017) on the data practices of temporary laborers in the South Pacific, and Lämmerhirt (2021) on citizen COVID-19 data donations. To clarify, the contributions to the present volume explore the data practices not just of professional practitioners, but also by and in relation to citizens in everyday situations. They also strive to combine situated understandings of data, as proposed by (critical) data studies scholars (e.g., Dencik 2019; Kennedy 2018), with a *greater* praxeological sensitivity previously identified by Marres. This brings with it several theoretical and methodological questions regarding our understanding of *situations* and *practices* that we will address in the following.

The praxeological approach to data suggested in this volume provides critical resources against the abstracting tendencies of datafication discussed above. Praxeological work may serve to empirically highlight and deconstruct how data are manufactured in practice, as well as inquire into the imaginaries around those same data practices (see Bucher 2017). As such, a praxeology of data is well-suited to study processes of datafication not from the God's eye perspective of a distanced and neutral theoretical observer (Haraway 1988, see the discussion in Rettberg 2020), but by following the trail of situated practices involved in processes of datafication.

Labeling the approach of the present volume as *praxeological* carries with it a number of premises and assumptions that are worth detailing. One such premise is the focus on studying the *ethnomethods* of specific communities of practice. This builds on the insight by Harold Garfinkel "that the activities whereby members produce and manage settings of organized everyday affairs are identical with members' procedures for making those settings 'account-able'" (Garfinkel 1967, 1). Social order, understood from an ethnomethodological perspective, is "assumed to display a mundane intelligibility of its own, prior to and independently of its scholarly treatment" (Sormani 2019, 3).

The effect of this focus on the accountability of social practices is twofold. Firstly, it pays attention to the inherent reflexivity of everyday practices, thus rejecting sociological understandings of practice that consider it unconscious, quasi-automatic, or unintelligible on behalf of the practitioner. Then secondly, such a focus places an emphasis on the specific methods of documentation (textual, audiovisual etc.) employed by members of a social group, such that they make their own activities "visibly-rational-and-reportable-for-all-practical-purposes" (Garfinkel 1967, vii). This in-built sensitivity to processes of mediation within ethnomethodology – made explicit in media formats like files, records, graphs, or audiovisual recordings – is further emphasized in a praxeological approach to data, as novel techniques, such as formatting practices (see Jancovic, Volmar, and Schneider 2020), are developed to enable the datafication of social practices.

Thus, a praxeological approach to data aims to avoid any preconceived notion of what a practice is, could, or should be, in relevant settings. Whilst proponents of ANT insist that one must "follow the actors themselves" (Latour 2005, 12), a praxeological approach, arguably, insists that one must "follow the action" instead (Boersma 2020, 665). In this formulation, the overriding interest is in articulating how rote actions – say, of registering people claiming asylum – become iterative practices. Or how new data practices are generated by technological developments in data collection, storage, and aggregation. Whilst social actors are, of course, critical to understanding how data practices are performed, it is the data *practices* as things-in-themselves, rather than the data *practitioners* that are the principal focus.

A praxeological account, therefore, has to begin its inquiry into data practices from traces of *observable* phenomena, not from general theoretical accounts or abstractions such as the notions of "data," "platform," or "society." Here, observable phenomena might well be indeterminate, ambiguous, or open-ended. Indeed, such phenomena might have multiple, attachable meanings and interpretations with its relevance only temporarily defined and negotiated. An example of such situational meaning-making is provided by Garfinkel in the distinction between actuarial and contractual uses of patient folders in a hospital setting (Garfinkel 1967, 197–207; Paßmann and Gerlitz 2014). In this, data are themselves situational, with meaning derived from the very practices into which they are enrolled.

With an interest in observability, praxeological approaches offer opportunities for various kinds of *ethnographic inquiry*. Whilst there are identifiable differences between historical/biographical forms of ethnography (in which observability of phenomena is somewhat difficult), traditional forms of embedded ethnography, and "shallower" forms of ethnography, there is nonetheless a shared interest in the study of communities of practice. Moreover, digital forms of ethnographic inquiry (like Kozinets' [2019] netnography or Pink et al.'s [2015] digital ethnography) recalibrate this shared interest as such communities of practice are distributed, yet infrastructurally stabilized through digital media and technologies. Likewise, digital methods approaches (Rogers 2013) consider the role that the medium itself has on how these communities are composed, rather than merely seeing them as incidental. Here, the project of a *data* ethnography – discussed in a moderated discussion in this volume – remains in its infancy, but necessarily draws on these traditions. In foregrounding the role of data in the ethnographic work undertaken, one is attentive to how it allows people to account for their activities, which, as this volume shows, can range from working in a hospital to driving a car. Rather than constituting an actor in its own right, however, data can be used to trace connections between and across sites in the fashion of a "multi-sited ethnography," that is "designed around chains, paths, threads, conjunctions, or juxtapositions of locations in which the ethnographer establishes some form of literal, physical presence, with an explicit, posited logic of association or connection among sites that in fact defines the argument of the ethnography" (Marcus 1995, 105). Following the data, thus, turns into a mode of field site construction, and necessarily complements the praxeological focus on observable actions and behavior.

Ethnographers of all persuasions do not only produce accounts of *other people's* data practices, but generate *their own data* while doing so. Data ethnographers thus take data practices to be their *object* of inquiry when investigating phenomena of datafication, whilst also employing data practices as a *method* through which to make sense of the scrutinized phenomena. Such data may, of course, be produced in concert with practitioners in the field, from hospital administrators to software engineers. Not limited to hand-written field notes or audio(visual) recordings, data may also take the form of relational database entries, data visualizations, outputs of automated tracking tools, or even experimental algorithms uploaded to software repositories. The result is a common overlap of research object and method of study that makes it necessary to explicate and reflect on the data practices employed on both sides. In any case, employing ethnographic methods to account for data practices brings with it a set of *new challenges* that warrant attention.

The first challenge concerns what a practice *actually is*. While a praxeology of data can build upon the intuitions of ANT and actor-media theory (Thielmann, Schüttpelz, and Gendolla 2013) to recognize the stakes of non-human actors in (re)assembling the social (Latour 2005), it is by no means a settled question how human and non-human agency is distributed in data practice(s). To what extent might semi-autonomous actors like robots, drones, and algorithms – and also less evocative elements such as dashboard dials and indicators – be said to engage in practices? This becomes all the more relevant when data processing happens without human oversight, e.g., within the layered architecture of a neural network-based machine learning algorithm (Hansen 2020). The question for a praxeology of data might, therefore, be in how the availability of data relates to and *intervenes in* processes of automation. A common way to talk about automation is framing it as *destroying* certain practices, most notably in how "robots" are framed as "taking jobs." Yet a praxeological approach might instead argue that practices are co-operative, but that certain kinds of automation may lead to categorical shifts in the *type* or *form* of data-related practices. For example, in how automation leads to the rise of new kinds of algorithmic "supervision" rendering decision-making procedures "opaque," as Annelie Pentenrieder's contribution to this volume on delivery couriers explores.

A second major challenge relates to the extent to which data-related phenomena are observable *in principle*. Whilst social practices do, in general, transcend the spatiotemporal confines of any given situation, data practices are inherently *trans-situational*. Only in particularly rare cases are data produced, gathered, archived, viewed, analyzed, or presented in one isolated situation, or a single location. Instead, one of the defining features of data is their capacity to detach from their original contexts and to be remotely processed (Leonelli and Tempini 2020). The development of a praxeology of data therefore requires the further refinement of multi-situated methods to describe data practices both *in situ* and *across situations*, which is required for the study of in our media landscape prominent highly distributed data infrastructures such as app ecosystems (Dieter et al. 2019; 2021) and the sensory media apps build and operate on (Chao et al. forthcoming). Such an approach is in line with George Marcus's proposal to not merely multiply studies of geographically bounded "sites" (Marcus 1995) but to draw attention to the environments conditioning people's circumstances of action (Paßmann and Schubert 2021). In opposition to traditional ethnographic approaches that focus their attention on "local" situations and processes of embodied inter-corporeality (i.e. forms of "co-presence"), a praxeology of data needs to deal with scalar and temporal *mediations* between situations. That is, to take into account what may *precede* any given situation and what it is prone to *develop into* (Gießmann and Röhl 2019; cf. also Goodwin 2018).

A third challenge is the extent to which data-related phenomena are *accessible*. While many media practices can be documented using established observational methods, data practices require *additional methods* that are sensitive to the dimension of background cooperation or partially autonomous processes characterizing contemporary digital culture. For example, gaining access to data processed on proprietary platforms, through closed ecosystems, or by sensor-equipped devices is often difficult if not impossible. This problem is usually framed around the notion of the black box these technologies present to both certain types of practitioners as well as researchers (Carabantes 2020; Latour 1999; Pinch 1992). It is also not entirely clear what access to data actually implies, and if, for example, the data practice of scraping social media data via an API already constitutes a privileged form of access in itself. The black box of digital media technologies is not a purely technical phenomenon either: it relates as well to the skill sets and expertise required by researchers to make sense of the investigated phenomena, thus necessitating the forming of inter- and transdisciplinary research teams. Data ethnographers need not – and from a certain perspective, should not – be data scientists themselves, but they are encouraged to develop inventive methods (Lury and Wakeford 2012) and experimental setups, such as breaching experiments (Rafalovich 2006), to make sense of the black boxes they encounter.

In summary, then, interrogating datafication from a praxeological perspective requires grappling with manifold challenges: from offering forms of critique that take seriously the societal implications of datafication (i.e. abstraction, automation, decontextualization, re-situation etc.), to developing creative methodologies that tackle the closed ecosystems common to contemporary media (e.g., data ethnography, interdisciplinarity, experimentation etc.). In the following, we provide a brief overview of the subsequent sections and chapters of the book, considering how they collectively interrogate datafication through studies of the data practices found in various social and cultural settings.

Chapter overview

The volume consists of four sections that discuss the history of data practices, the possibilities and challenges of data ethnography, the entanglements of data and care practices as well as the relationship between data practices and mobility.

Section 1 opens with a contribution by Kyle Stine about the history of film as a data medium. Stine makes the case that before the advent of electronic digital computing, motion picture film served as the first universal data medium in its ability to translate between image, sound, text, and machine movement. Within the period of "the long 1920s," which Colin Koopman has cited as the genealogical root of the "informational person," film acted as a point of coordination for different data sources, a medium of coherence and universality, that mirrored the informational person as a common body for disparate data being collected in the forms and checkboxes of birth and death certificates, identification papers, medical records, racialized credit information, and police files (Koopman 2019). Across a variety of fields, inventors sought to use film as a more economical and efficient way to search and retrieve data from these growing stacks of information. Ultimately, film would not live up to the visions its advocates had of it, as microfilm search-and-retrieval systems such as Emanuel Goldberg's Statistical Machine and Vannevar Bush's Memex would find narrow applications or not be implemented at all. But the ambitions to use film for data-processing purposes provides unique insight into the problems that digital computing would be mobilized to solve, and the cultural and political values that drove such efforts.

In the second contribution to the first section, Liam Cole Young explores the emergence of Hollywood box office charts, arguing such charts offer a crucial step in the genealogy of contemporary cloud-based forms of tracking, prediction, and decision-making as found in recommendation and newsfeed algorithms. Young adopts a praxeological approach that emphasizes where and how box office data came to be aggregated and displayed, as well as some of the motivations that led people to *datafy* the production, distribution, and reception of cinema. This approach allows us to see how box office data was put to work, how it came to reconfigure chains of decision-making and resource distribution within the film industry. Box office charts and other *datafied* forms rarely appear in histories of culture, while accounts of contemporary data analytics typically begin only with the digital computer. Young's aims are thus historical and deflationary, to show how deeper histories of counting culture prefigure and anticipate today's Big Data, and to weave practices of accounting into stories about cultural industries that can complement the usual emphasis on production and consumption.

Section 2 consists of two contributions that explore the lens, advantages and possibility of data ethnography as a praxeologically informed approach to the study of the datafication of everyday, social life. The first contribution features a moderated conversation about the challenges that studying data practices brings about, its epistemological consequences, and the methodological prerequisites this enterprise asks for. The second chapter in this section is an empirical experiment in data ethnography, combining a politico-technological perspective on institutional data practices with the perspective of the subject who is confronted with datafied systems of state governance.

In the moderated conversation on "Doing Data Ethnography," Daniela van Geenen and Danny Lämmerhirt discuss with Emma Garnett, Minna Ruckenstein, Tommaso Venturini, and Malte Ziewitz, four scholars with backgrounds in anthropology, sociology and STS who study data-intensive phenomena, how ethnography informs their research *with* and *on* data. Probing the question of what data ethnographic research practices could look like, the conversation addresses several pertinent questions of a social study of data: How do ethnographic sensibilities create unexpected perspectives on data? How can ethnographic studies account for distributed data practices? How should one methodologically attune to the study of data practices? What kinds of collaboration and positions may ethnographers take with and about data? The chapter emphasizes the importance of ethnographic sensibilities to consider and reflect on one's own entanglement with involved devices, data, and practices. An important aspect of data ethnographic research is the ability to situate data, to inquire into and document people's own understandings of data, and to provide reflexive accounts of one's own research practices. As such, data ethnography may foster and furnish important praxeological sensibilities in response to dominant data science paradigms. Data ethnographic approaches may open up spaces for dialog and reflection on the ideologies and values underpinning data collections, the often messy practices involved in the construction and use of data, and the surprising perspectives, unexpected questions, and insights one might gain from situating data.

Building on these ethnographic sensibilities, the empirical chapter in section 2 by Araa Al Jaramani, Sandra Ponzanesi, and Gerwin van Schie shows the relevance of the immigration procedure of the Dutch Immigration and Naturalization Service (IND) from the perspective of the data subjects forced to relate to the ways in which the IND's datafied bureaucratic system governs the asylum procedure. The authors combine a top-down perspective (data system) with a bottom-up perspective (data subjects) integrating an analysis of data and information about five Syrian refugee women who have faced the IND's opaque decision-making process for granting (or denying) the right to asylum. In this chapter, the women speak back to the system, producing alternative knowledge and representations to the dominant and mainstream stories of migration and integration in the Netherlands. The authors have backgrounds in (digital) media, post-colonial and critical data studies, gender studies, and accounts by the first author are partially auto-ethnographic.

The contributions in *section 3* unpack the entanglements of data and care practices from different perspectives. Here, data as well as data-intensive technologies appear not only as increasingly important means of care in contexts such as healthcare and elderly care, but as genuine objects of care. The contributors show the various ways in which data requires care and is cared for when it is recorded, interpreted, used, shared, archived, or reused. From a critical perspective this focus on data as an object of care confronts the imaginaries of big and open data with the realities of data practices that are contingent, full of frictions, and laborious. As a mode of care this work is far from automatic but is involved, engaged, attentive, and reflexive. Such a perspective contributes to a more nuanced analysis of data power as much as it feeds back into the design of technologies that support reflexive, caring data practices and the careful design of future technologies.

In their paper *Everyday Curation* Kate Weiner et al. discuss practices of self-monitoring of health-related data such as blood pressure or body weight. Drawing on an interview study the authors propose to conceptualize the data practices of individuals as curatorial practices. These practices of data curation entail *discerning work*, for example in the selection of relevant readings to become part of data records. As a result, such records only consist of *partial data* according to the authors that is recorded, interpreted, and circulated by engaged individuals rather than by disengaged (quasi-)machinic processes. The contribution of Claudia Müller and David Struzek is concerned with user practices of data-intensive technologies as well. With their background in Socio-Informatics they ask how future users can be involved in the development and design of digital technologies that aim to support them. This is of special importance for user groups that have little or no knowledge, expertise, or affinity to digital technologies. Building upon experiences gained from a participatory project for developing sensor technologies *with* and *for* older adults in rural areas the authors argue for grounding technology development in everyday practices and user needs. Contrary to many high-tech imaginaries the digital revolution begins here with off-the shelf devices and the participatory design of mutual practices.

Within academic research the Open Science movement has promoted the idea of openly accessibly and reusable research data. While this development was mainly driven by the natural sciences and engineering it by now affects academic research at large. Against this background Wolfgang Kraus and Igor Eberhard discuss in their contribution the challenges they face in setting up the *Ethnographic Data Archive* (EDA) at the University of Vienna. Struggling with both a reductionist understanding of research that underlies many discussions on Open Research Data, and the skepticism of many qualitative researchers, the EDA aims at developing best-practice models for archiving ethnographic research data that are sensitive to the specificities of ethnographic research and the dialogic creation of ethnographic data. The contribution by Gaia Mosconi et al. is similarly concerned with qualitative and ethnographic research data. Drawing on insights from discourses on data storytelling and empirical research of data practices of ethnographic researchers, the authors propose the design concept of Data Story that aims at supporting researchers to select relevant data snippets and to enhance them with contextual information in narrative structure. The creation of such Data

Stories is conceptualized as a form of selective care that increases the usefulness as well as reusability of qualitative data by interweaving formal descriptions with informal narratives of data in a structured, yet adaptable process.

Section 4 focuses on the relationship of data practices and mobility. Generating "real-time" data about people and things on the move and thus in dynamic, real-world settings is one of the main trajectories of datafication, and is increasingly taken up in inventive and (mobile) digital methods to study social phenomena in motion. The contributions in this section focus on urban pedestrian mobility (O'Grady), the datafication of driving (Hind), and the logistics sector (Pentenrieder). Together, they explore the ways in which the tracking and monitoring of behaviors, actions, and experiences produces data, but also how data acts to inform and modulate everyday practices.

Nathaniel O'Grady offers an analysis of a public WiFi infrastructure that has the capacity to generate what he calls "affective atmospheres," actively shaping public practices in New York. LinkNYC, the operator of 10ft high WiFi kiosks found across the city, offers the opportunity for advertisers who use the kiosks' 55inch screens for their adverts, to understand – and target – the daily journeys of prospective customers. Here, we see the possibility of such an infrastructure as firstly being able to triangulate pedestrian movements as people drop in and out of each kiosk's WiFi range, courtesy of their mobile phone's connectivity. But, secondly, in capturing these user journeys, how their embodied and affective *experience* of walking through the city – i.e. principally what they might *see* – can be *modulated* in relation to the specific encounters that the LinkNYC kiosks offer.

Sam Hind likewise considers how a range of driving-related phenomena are turned into data, and how particular systems and interfaces within the car are designed to re-structure relations between driver and vehicle, and in the process transforming the driving experience into a datafied one. What is interesting here is the extent to which new driving experiences emerge when automobiles become platformized hubs of multiple data streams. Hind outlines the various ways geodata is transformed into navigational data, i.e. data informing and remodeling the navigational process on a *turn-by-turn* basis, and vehicle data is surfaced as driving data, best exemplified by how car dashboard interfaces are converging both spatially and operationally. The contribution delves deeper into two strategies employed by car manufacturers to hook drivers in: a process of "representational transparency" meant to smooth the navigational experience, and the offering of forms of "customizable control" designed to personalize the driving experience.

Annelie Pentenrieder considers how software modulates the work of drivers in the logistics sector, and specifically, how route planners and digital maps help coordinate trips, distributes tasks, and control the execution of work-related activities. Here, Pentenrieder considers how such software generates *algorithmic opacities* such that drivers are unable to see or challenge the distribution of, for example, incoming delivery orders assigned to them. Unlike human supervisors, as Pentenrieder explores, algorithms cannot be questioned, leaving drivers oblivious to the reasoning behind decisions. Accordingly, the contribution calls for a re-examination of algorithmic opacities from the perspective of the user. In short, to make sense of the *user experience*, from long-distance lorry drivers to food couriers. In this (re)focus on the everyday interactions these users have with software that governs their daily activities, we gain a deeper understanding of how they develop strategies and deploy tricks to make sense of the logics of software as they are deployed.

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

References

Agre, Philip. 1994. "Surveillance and Capture: Two Models of Privacy." *The Information Society: An International Journal* 10 (2): 101–127.

Algorithm Watch. 2020. "Automated Decision-Making Systems in the COVID-19 Pandemic: A European Perspective." Automating Society Report 2020. Accessed January 18, 2022. https://algorithmwatch.org/en/wp-content/uploads/2020/08/ADM-systems-in-the-Covid-19-pandemic-Report-by-AW-BSt-Sept-2020.pdf.

Andrejevic, Mark. 2007. "Surveillance in the Digital Enclosure." *The Communication Review* 10 (4): 295–317. Accessed January 18, 2022. https://doi.org/10.1080/10714420701715365.

Aronova, Elena, Christine von Oertzen, and David Sepkoski. 2017. "Introduction: Historicizing Big Data." *Osiris* 32 (1): 1–17. Accessed January 18, 2022. https://doi.org/10.1086/693399.

Bates, Jo, Yu-Wei Lin, and Paula Goodale. 2016. "Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows." *Big Data & Society* 3 (2): 1–12. Accessed January 18, 2022. https://doi.org/10.1177%2F2053951716654502.

Boersma, Asher. 2020. "Follow the Action: Understanding the Conflicting Temporalities of Ships, River, Authorities and Family Through Distributed Ethnography." *Mobilities* 15 (5): 661–676. Accessed January 18, 2022. https://doi.org/10.1080/17450101.2020.1823135.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–79. Accessed January 18, 2022. https://doi.org/10.1080/1369118X.2012.678878.

Bucher, Taina. 2017. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20 (1): 30–44. Accessed January 18, 2022. https://doi.org/10.1080/1369118X.2016.1154086.

Carabantes, Manuel. 2020. "Black-Box Artificial Intelligence: An Epistemological and Critical Analysis." *AI & Society* 35 (2): 309–17. Accessed January 18, 2022. https://doi.org/10.1007/s00146-019-00888-w.

Correll, Michael, and Heather Froehlich. 2020. "The Spectacular Dashboard." *mcorrell.medium*. Accessed January 18, 2022. https://mcorrell.medium.com/the-spectacular-dashboard-dcb190ed8529.

Chao, Jason, Geenen, Daniela van, Gerlitz, Carolin, and Fernando van der Vlist. Forthcoming. "Digital Methods for Sensory Media Research." *Convergence.*

Christin, Angèle. 2017. "Algorithms in Practice: Comparing Web Journalism and Criminal Justice." *Big Data & Society* 4 (2): 1–14. Accessed January 18, 2022. https://doi.org/10.1177%2F2053951717718855.

Couldry, Nick. 2020. "Recovering Critique in an Age of Datafication." *New Media & Society* 22 (7): 1135–51. Accessed January 18, 2022. https://doi.org/10.1177%2F1461444820912536.

Crooks, Roderic, and Morgan Currie. 2021. "Numbers Will Not Save Us: Agonistic Data Practices." *The Information Society* 37 (4): 201–13. Accessed January 18, 2022. https://doi.org/10.1080/01972243.2021.1920081.

Dalton, Craig, and Jim Thatcher. 2014. "What Does A Critical Data Studies Look Like, And Why Do We Care?" *Society and Space*. Accessed January 18, 2022. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care.

Dalton, Craig M, Linnet Taylor, and Jim Thatcher. 2016. "Critical Data Studies: A Dialog on Data and Space." *Big Data & Society* 3 (1): 1–9. Accessed January 18, 2022. https://doi.org/10.1177%2F2053951716648346.

Day, Sophie, Celia Lury, and Nina Wakeford. 2014. "Number Ecologies: Numbers and Numbering Practices." *Distinktion: Journal of Social Theory* 15 (2): 123–154. Accessed January 18, 2022. https://doi.org/10.1080/1600910X.2014.923011.

de Chadarevian, Soraya, and Theodore M. Porter. 2018. "Introduction: Scrutinizing the Data World." *Historical Studies in the Natural Sciences* 48 (5): 549–556. Accessed January 18, 2022. https://doi.org/10.1525/hsns.2018.48.5.549.

Dencik, Lina. 2019. "Situating Practices in Datafication – from Above and Below." In *Citizen Media and Practice: Currents, Connections, Challenges*, edited by Hilde C. Stephensen and Emiliano Treré, 243–55. London: Routledge.

Dieter, Michael, Carolin Gerlitz, Anne Helmond, Nathaniel Tkacz, Fernando N. van der Vlist, and Esther Weltevrede. 2019. "Multi-Situated App Studies: Methods and Propositions." *Social Media + Society* 5 (2): 1–15. Accessed January 18, 2022. https://doi.org/10.1177%2F2056305119846486.s

Dieter, Michael, Anne Helmond, Nathaniel Tkacz, Fernando van der Vlist, and Esther Weltevrede. 2021. "Pandemic Platform Governance: Mapping the Global Ecosystem of COVID-19 Response Apps." *Internet Policy Review* 10 (3): 1–28. Accessed January 18, 2022. https://doi.org/10.14763/2021.3.1568.

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

Dijck, Jose van. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12 (2): 197–208. Accessed January 18, 2022. https://doi.org/10.24908/ss.v12i2.4776.

Dijck, José van, Thomas Poell, and Martijn de Waal. 2018. *The Platform Society*. Oxford: Oxford University Press.

Es, Karin van, and Eef Masson. 2018. "Big Data Histories: An Introduction." *Tijdschrift voor Mediageschiedenis* 21 (2): 1–6. Accessed January 18, 2022. http://doi.org/10.18146/2213-7653.2018.363.

European Commission. 2020. "Proposal for a Regulation on European Data Governance (Data Governance Act)." *European Commission*. Accessed January 18, 2022. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767.

Fourcade, Marion, and Kieran Healy. 2017. "Seeing like a Market." *Socio-Economic Review* 15 (1): 9–29. Accessed January 18, 2022. https://doi.org/10.1093/ser/mww033.

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

Gerlitz, Carolin. 2016. "What Counts? Reflections on the Multivalence of Social Media Data." *Digital Culture & Society* 2 (2): 19–38. Accessed January 18, 2022. https://doi.org/10.14361/dcs-2016-0203.

Gerlitz, Carolin, and Anne Helmond. 2013. "The Like Economy: Social Buttons and the Data-intensive Web." *New Media & Society* 15 (8): 1348–1365. Accessed January 18, 2022. https://doi.org/10.1177%2F1461444812472322.

Gerlitz, Carolin, and Celia Lury. 2014. "Social Media and Self-evaluating Assemblages: On Numbers, Orderings and Values." *Distinktion: Scandinavian Journal of Social Theory* 15 (2): 174–188. Accessed January 18, 2022. https://doi.org/10.1080/1600910X.2014.920267.

Gießmann, Sebastian. 2018. "Elemente einer Praxitheorie der Medien." *Zeitschrift für Medienwissenschaft* 19 (2): 95–109. Accessed January 19, 2022. https://doi.org/10.25969/mediarep/1228.

Gießmann, Sebastian, and Tobias Röhl. 2019. "Materialität der Kooperation zur Einleitung." In *Materialität der Kooperation*, edited by Sebastian Gießmann, Tobias Röhl, and Ronja Trischler, 3–32. Wiesbaden: Springer Fachmedien. Accessed January 19, 2022. https://doi.org/10.1007/978-3-658-20805-9_1.

Goodwin, Charles. 2018. *Co-Operative Action*. Cambridge: Cambridge University Press.

Hansen, Kristian Bondo. 2020. "The Virtue of Simplicity: On Machine Learning Models in Algorithmic Trading." *Big Data & Society* 7 (1): 1–14. Accessed January 19, 2022. https://doi.org/10.1177/2053951720926558.

Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99. Accessed January 19, 2022. https://doi.org/10.2307/3178066.

Hobbis, Geoffrey. 2017. "The MicroSDs of Solomon Islands: An Offline Remittance Economy of Digital Multi-Media." *Digital Culture & Society* 3 (2): 21–39. Accessed January 19, 2022. https://doi.org/10.14361/dcs-2017-0203.

Iliadis, Andrew, and Federica Russo. 2016. "Critical Data Studies: An Introduction." *Big Data & Society* 3 (2): 1–7. Accessed January 19, 2022. https://doi.org/10.1177/2053951716674238.

Isin, Engin, and Evelyn Ruppert. 2020. "The Birth of Sensory Power: How a Pandemic Made It Visible?" *Big Data & Society* 7 (2): 1–15. Accessed January 19, 2022. https://doi.org/10.1177/2053951720969208.

Jancovic, Marek, Axel Volmar, and Alexandra Schneider, eds. 2020. *Format Matters: Standards, Practices, and Politics in Media Cultures*. Lüneburg: meson press.

Kennedy, Helen. 2018. "Living with Data: Aligning Data Studies and Data Activism Through a Focus on Everyday Experiences of Datafication." *Krisis: Journal for Contemporary Philosophy* 1: 18–30.

Kennedy, Helen, Susan Oman, Mark Taylor, Jo Bates, and Robin Steedman. 2020. "Public Understanding and Perceptions of Data Practices: A Review of Existing Research." *Livingwithdata.org*. Accessed January 19, 2022. https://livingwithdata.org/project/wp-content/uploads/2020/05/living-with-data-2020-review-of-existing-research.pdf.

Kerssens, Niels, and José van Dijck. 2021. "The Platformization of Primary Education in The Netherlands." *Learning, edia and Technology* 46 (3): 250–63. Accessed January 19, 2022. https://doi.org/10.1080/17439884.2021.1876725.

Kitchin, Rob. 2014a. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1): 1–12. Accessed January 19, 2022. https://doi.org/10.1177/2053951714528481.

Kitchin, Rob. 2014b. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oaks, CA: Sage.

Kitchin, Rob, and Tracey Lauriault. 2018. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." In *Thinking Big Data in Geography: New Regimes, New Research*, edited by Jim Thatcher, Josef Eckert, and Andrew Shears. Lincoln, NE: University of Nebraska Press.

Knorr-Cetina, Karin. 2009. “The Synthetic Situation: Interactionism for a Global World.” *Symbolic Interaction* 32 (1): 61–87. Accessed January 19, 2022. https://doi.org/10.1525/si.2009.32.1.61.

Koopman, Colin. 2019. *How We Became Our Data: A Genealogy of the Informational Person*. Chicago, IL: University of Chicago Press.

Kozinets, Robert. 2019. *Netnography: The Essential Guide to Qualitative Social Media Research*. Thousand Oaks, CA: Sage.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Lämmerhirt, Danny. 2021. “Health Data Donations as Cooperative Practice.” *SFB1187 Media of Cooperation Working Paper Series* 18: 30–34. Accessed January 19, 2022. https://doi.org/10.25819/ubsi/9948.

Langlois, Ganaele, Greg Elmer, Fenwick McKelvey, and Zachary Devereaux. 2009. “Networked Publics: The Double Articulation of Code and Politics on Facebook.” *Canadian Journal of Communication* 34 (3): 415–434. Accessed January 19, 2022. https://doi.org/10.22230/cjc.2009v34n3a2114.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1999. *Pandora’s Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Leonelli, Sabina. 2021. “Data Science in Times of Pan(dem)ic.” *Harvard Data Science Review*, January. Accessed January 19, 2022. https://doi.org/10.1162/99608f92.fbb1bdd6.

Leonelli, Sabina, Robert P. Davey, Elizabeth Arnaud, Geraint Parry, and Ruth Bastow. 2017. “Data Management and Best Practice for Plant Science.” *Nature Plants* 3. Accessed January 19, 2022. https://doi.org/10.1038/nplants.2017.86

Leonelli, Sabina and Niccolò Tempini, eds. 2020. *Data Journeys in the Sciences*. Cham: Springer.

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

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

Marres, Noortje. 2012. “The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived.” *The Sociological Review* 60 (1): 139–65. Accessed January 19, 2022. https://doi.org/10.1111/j.1467-954X.2012.02121.x.

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

Mattern, Shannon. 2014. "Interfacing Urban Intelligence." *Places Journal*. Accessed January 19, 2022. https://placesjournal.org/article/interfacing-urban-intelligence/.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.

Mejias, Ulises A., and Nick Couldry. 2019. "Datafication." *Internet Policy Review* 8 (4): 1–10. Accessed January 19, 2022. https://doi.org/10.14763/2019.4.1428.

Milan, Stefania, and Emiliano Treré. 2019. "Big Data from the South(s): Beyond Data Universalism." *Television & New Media* 20 (4): 319–35. Accessed January 19, 2022. https://doi.org/10.1177/1527476419837739.

Milan, Stefania, Emiliano Treré, and Silvia Masiero. 2021. *COVID-19 from the Margins. Pandemic Invisibilities, Policies and Resistance in the Datafied Society*. Amsterdam: Institute of Network Cultures.

Neff, Gina, 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): 85–97. Accessed January 19, 2022. https://doi.org/10.1089/big.2016.0050.

Paßmann, Johannes, and Carolin Gerlitz. 2014. "'Good' Platform-Political Reasons for 'bad' Platform-Data. Zur Sozio-Technischen Geschichte der Plattformaktivitäten Fav, Retweet und Like." *Mediale Kontrolle* 3 (1): 1–40.

Paßmann, Johannes, and Cornelius Schubert. 2021. "Technografie als Methode der Social-Media- Forschung." In *Handbuch Diskurse Digital*, edited by Konstanze Marx, Simon Meier-Vieracker, and Friedemann Vogel. Berlin: De Gruyter. https://doi.org/10.13140/RG.2.2.22738.50883.

Pinch, Trevor J. 1992. "Opening Black Boxes: Science, Technology and Society." *Social Studies of Science* 22 (3): 487–510. Accessed January 19, 2022. https://doi.org/10.1177/030631279202200 3003.

Pink, Sarah, Heather A. Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi, eds. 2015. *Digital Ethnography: Principles and Practice*. Thousand Oaks, CA: Sage.

Pink, Sarah, Shanti Sumartojo, Deborah Lupton, and Christine Heyes La Bond. 2017. "Mundane Data: The Routines, Contingencies and Accomplishments of Digital Living." *Big Data & Society* 4 (1): 1–12. Accessed January 19, 2022. https://doi.org/10.1177/2053951717700924.

Poell, Thomas, David Nieborg, and José van Dijck. 2019. "Platformisation." *Internet Policy Review* 8 (4): 1–13. Accessed January 19, 2022. https://policyreview.info/concepts/platformisation.

Radin, Joanna. 2017. "'Digital Natives': How Medical and Indigenous Histories Matter for Big Data." *Osiris* 32 (1): 43–64. Accessed January 19, 2022. https://doi.org/10.1086/693853.

Rafalovich, Adam. 2006. "Making Sociology Relevant: The Assignment and Application of Breaching Experiments." *Teaching Sociology* 34 (2): 156–63. Accessed January 19, 2022. https://doi.org/10.1177/0092055X0603400206.

Rettberg, Jill Walker. 2020. "Situated Data Analysis: A New Method for Analysing Encoded Power Relationships in Social Media Platforms and Apps." *Humanities and Social Sciences Communications* 7 (1): 1–13. Accessed January 19, 2022. https://doi.org/10.1057/s41599-020-0495-3.

Rheinberger, Hans-Jörg. 2011. "Infra-Experimentality: From Traces to Data, from Data to Patterning Facts." *History of Science* 49 (3): 337–348. Accessed January 19, 2022. https://doi.org/10.1177/007327531104900306.

Rogers, Richard. 2013. *Digital Methods*. Cambridge, MA: MIT Press.

Ruckenstein, Minna. 2014. "Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles." *Societies* 4 (1): 68–84. Accessed January 19, 2022. https://doi.org/10.3390/soc4010068.

Ruppert, Evelyn, Engin Isin, and Didier Bigo. 2017. "Data Politics." *Big Data & Society* 4 (2): 1–7. Accessed January 19, 2022. https://doi.org/10.1177/2053951717717749.

Ruppert, Evelyn, and Stephan Scheel. 2021. *Data Practices: Making Up a European People*. London: Goldsmiths Press.

Shapiro, Aaron. 2020. *Design, Control, Predict: Logistical Governance in the Smart City.* Minneapolis, MN: University of Minnesota Press.

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

Schüttpelz, Erhard, and Christian Meyer. 2017. "Ein Glossar zur Praxistheorie." *Navigationen* 17 (1): 155–163.

Smith, Gavin JD. 2018. "Data Doxa: The Affective Consequences of Data Practices." *Big Data & Society* 5 (1): 1–15. Accessed January 19, 2022. https://doi.org/10.1177/2053951717751551.

Sormani, Philippe. 2019. "Ethnomethodological Analysis." In *SAGE Research Methods Foundations*, edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshuag, and Richard A. Williams. Thousand Oaks, CA: Sage. Accessed January 19, 2022. https://doi.org/10.4135/9781526421036788330.

Southerton, Clare. 2020. "Datafication." In *Encyclopedia of Big Data*, edited by Laurie A. Schintler and Connie L. McNeely, 1–4. Cham: Springer. Accessed January 19, 2022. https://doi.org/10.1007/978-3-319-32001-4_332-1.

The Economist. 2021. "Tracking the Coronavirus across Europe," 28 December 2021. Accessed January 19, 2022. https://www.economist.com/graphic-detail/tracking-coronavirus-across-europe.

Thielmann, Tristan. 2019. "Sensormedien: Eine medien- und praxistheoretische Annäherung." *SFB1187 Media of Cooperation Working Paper Series* 9: 1–12.

Thielmann, Tristan, Erhard Schüttpelz, and Peter Gendolla, eds. 2013. *Akteur-Medien-Theorie*. Bielefeld: Transcript.

Tkacz, Nathaniel, Mário Henrique da Mata Martins, João Porto de Albuquerque, Flávio Horita, and Giovanni Dolif Neto. 2021. "Data Diaries: A Situated Approach to the Study of Data." *Big Data & Society* 8 (1): 1–16. Accessed January 19, 2022. https://doi.org/10.1177/2053951721996036.

Williamson, Ben. 2016. "Calculating Children in the Dataveillance School: Personal and Learning Analytics." In *Surveillance Futures: Social and Ethical Implications of New Technologies for Children and Young People*, edited by Emmeline Taylor and Tonya Roobey, 50–66. New York: Routledge.

World Health Organization. 2020. "Understanding the Infodemic and Misinformation in the Fight Against COVID-19." *WHO*. Accessed January 19, 2022. https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic_eng.pdf

I: Cultural Histories of Data

Film as the First Universal Data Medium

Kyle Stine

In describing the single-memory, stored-program computer architecture that would come to be known by his name, John von Neumann included among the suitable media for storing instructions motion picture film:

> These instructions must be given in some form which the device can sense: Punched into a system of punchcards or on teletype tape, magnetically impressed on steel tape or wire, *photographically impressed on motion picture film*, wired into one or more fixed or exchangeable plugboards – this list being by no means necessarily complete. (von Neumann [1945] 1993, 33; Kirschenbaum 2008, 27; emphasis added)

Von Neumann's 1945 report on the EDVAC came at the tail end of film's efficacy in computing. Two years later in his "Lecture on the Automatic Computing Engine," Alan Turing would strike down the idea of using film for its being uneconomical even before ruling it out because it could not provide storage that was erasable (Turing [1947] 2004, 380). Although Turing noted another technology used in visual reproduction, the cathode-ray tube, as the likely "ultimate solution," neither motion picture film nor the television component would provide storage for the von Neumann architecture, as that role would be fulfilled by magnetic cores and later silicon (see relatedly Chun 2008; Gaboury 2021). But to disregard film as a dead-end in the history of computing is to miss the many influences that developments in photography and film have had on the advancement of electronics and semiconductors (Stine 2019). Moreover, from a media-archaeological perspective, there is much to be gained from excavating early-twentieth-century instances of film being applied to problems in data processing and storage because a case can be made that film was in fact the first universal medium.

Before digital computers and graphical user interfaces, film was the most readily available means of storing image, sound, text, and data and coordinating between them. Film provided a computing and control medium, means of document storage and retrieval, and ways of interacting with data, in both experimental and established systems between the 1930s and 1960s, that would prefigure many of the functions of later computers and serve as inspiration for porting those functions from celluloid to silicon. The widespread availability of film, owing to the popular demand for motion pictures, made it an attractive medium for shoring up some of the inadequacies of its predecessor and rival, paper, which was used as a substrate for coding and printout in the form of paper cards and paper tape but was not as durable or amenable to applications with new electrical systems as was film (see Gitelman 2014). Film enabled the automatic transcoding between mechanical movement, light, and electricity, and through these the reproduction and translation of images, sounds, texts, and computational coding systems.

Film came to be used in a wide range of data practices in the period of transformation that Colin Koopman has called "the long 1920s," when census data, employment forms, medical and health records, and racialized credit information coalesced into an "informational personhood" that gave rise to the later growth of information theory with the writings of Claude Shannon and Norbert Wiener (Koopman 2019, x). Paper remained the primary means of data input, with forms and checkboxes serving as data-collection formats. But like the US national census records in the 1880s, these forms became so numerous that they exceeded effective means of filing and data processing, pushing efforts to automate practices of form reading. Herman Hollerith's solution to the census problem was, famously, to use punched cards capable of being tabulated by machine, inspired by the railway practice of the "punch photograph," where a train operator would use a punch card to indicate hair color, eye color, etc., to indicate that a passenger had paid (Hollerith 1971). Hollerith understood his census tabulator to allow for a kind of pictorial impression of mass population data. Inventors and scientists regarded motion picture film similarly as an ideal medium for comprehending the piles of paper data being collected in medicine, insurance, and policing. To the extent that it was implemented, film accelerated and expanded data processing, substituting data operations for some of the labor of data practices. Even in its failures, it is an important prefiguration of today's coordinated systems of universal digital data, and offers historical reference points for the cultural and political hazards of surveillance techniques.

In what follows, I offer a breakdown of the three major ways film was used as a data medium, which correspond to Friedrich Kittler's three media functions of storage, transmission, and data processing (2009, 30). Film was first and foremost a storage medium capable of reproduction, a feature noted variously by film theorists such as André Bazin ([1945] 1967), who saw in it a deep potential to preserve likeness against the passage of time, and information scientists such as Paul Otlet ([1901] 1990), who envisioned its pictorial detail, in the form of microfilm, as a means of preserving documents. Through its interaction with and influence on electrical technologies, such as photomultipliers, film was also integral to systems of data transmission, where through the process of transduction from light into electricity, film became the support for sending data through information systems. Lastly, film enabled data processing in early analog and digital computers, while also coming into use in early automation systems. As I will show, it is in acting as a coordinating point for these three media functions that film earns consideration as a proto-computational medium and a testing ground for automating twentieth-century data practices.

Durable, Flexible, and Transparent

The conditions for film's use in a number of different data practices were set at the medium's beginning. Durable, flexible, and transparent, film could withstand repeated mechanical action and receive the imprint of nearly any visual design. These properties were discovered, as with many inventions in the history of media, through a combination of accident and scientific experimentation. As Deac Rossell relates, the discovery of celluloid, created by soaking cotton or wood fibers in nitric acid and a suitable solvent, "led directly to the formation of the field of organic chemistry" and ushered in a range of new chemical products (1998, 63). The German-Swiss chemist Christian Friedrich Schönbein, in a letter to Michael Faraday in 1846, extolled the virtues of paper created through this process, saying, "I have of late also made a little chemical discovery which enables me to change very suddenly, very easily and very cheaply common paper in such a way as to render that substance exceedingly strong and entirely water proof" (emphasis in original; Faraday 1991, 477; as cited in Rossell 1998, 58). The addition of photosensitive emulsions to cellulose would take a different route initially, with collodion, or nitrocellulose gel, being used in wet-plate photography, such as in the inexpensive tintypes popularized during the American Civil War. It was only a small step, however, to propose that this material, which was used – dangerously, given its flammability – for durable objects such as dental plates, knife handles, and piano keys, could leave behind the glass plates of early photographic processes and become, as Daniel Spill presented to the London Photographic Society, "a flexible and structureless substitute for the glass negative supports" (as cited in Rossell 1998, 63). Liberating the photographic emulsion from rigid glass plates would lead to the introduction of roll film and roll-film holders, bringing together the flexible, durable medium of celluloid with a new economical form of storage and mechanism of transport. Storage and transport, in turn, would lay the foundation for a wide array of data practices in film.

In the early twentieth century, Paul Otlet, founding figure of information science and deviser of the Universal Decimal Classification, turned to what must be assumed to be cellulose acetate film, or safety film, as a "tough, stable, non-flammable" answer to the problem of preserving and making accessible libraries of books ([1901] 1990). "As early as 1906," Otlet later wrote, "we proposed that the book or documents generally should be given a new form, that of the miniature 'volumen' as follows: each page, element, or combination of pages is photographed directly on a 'frame' or film of the standard motion picture format" ([1925] 1990). The remediation of textual materials on film, Nanna Bonde Thylstrup (2019) has suggested, served both to preserve and to extend the reach of library collections, prefiguring in analog what would later take place in mass digitization. Emphasis in accounts of Otlet's work tends to fall, understandably, on the capacity of film to register textual and pictorial detail photographically, but equally important in Otlet's visionary essays were the mechanical means of transport and how the individual frames came together to "make up a microphotographic reel" ([1901] 1990). Were each image to be on a separate celluloid sheet, the user would have to reload the microfilm or microfiche reader with each page. A reel introduced in informational practice exactly what it introduced in popular cinema: movement. Helmut Müller-Sievers puts this in perspective in seeing "the film camera as a lathe that carves light onto film" (2001, 42), emphasizing the mechanical properties of the apparatus and understanding the film reel to function like any cylinder in a kinematic chain, bringing repeatable mechanical movement to an operation. This repeatable mechanical movement enabled the illusion of life in cinema, and it made possible new access to information on microfilm. To understand the full extent of film's influence on information science, however, requires an understanding of how the storage function on film related to its transmission function, and for that we need to pass through the history of cinema.

The Analog Principle and Photocells

The most information-oriented developments in microfilm concern the soundtrack more than the image track; or rather, they show that the image track is simply one filmic use case centered on lens-based picturing. Inventors from Eugene Lauste to Lee de Forest viewed motion picture film not simply as a means of figural representation but as a translational medium capable of carrying image information that might be converted into sound. The qualities that made film amenable to capturing scenes in photographic detail also made it well-suited for differentiating between elements of analog information. That several standards were developed for sound on film, namely the variable-area and variable-density methods, points to the openness of motion picture film to different informational encodings. The basis of these information tracks had as much to do with the method of reading the light passing through or blocked by the celluloid strip as with the strip itself. That is, the significance of film to data practices extends to the component technologies that it served to advance, the most important of which were photocells.

The first such light-reader, or photosensor, was discovered quite by accident when Willoughby Smith, working on the transatlantic cable in 1872, adopted selenium in hopes of making use of its very high electrical resistance at the English shore end of the line. Smith found that under certain conditions selenium bars showed extraordinary resistance, as high as 1,400 megohms, or as Alexander Graham Bell later put it in perspective, "a resistance equivalent to that which would be offered by a telegraph wire long enough to reach from the earth to the sun!" (1880, 132). But when the bars were exposed to light, the resistance dropped precipitously. Bell for his part would take up this feature, as he explained in a lecture to the Royal Society in May 1878, to devise a way of "hearing a shadow by means of interrupting the action of light upon selenium" (1880, 132). His dream was to develop wireless telephony, a way of sending the voice over great distances using light alone, but he quickly realized that sound could translate into any number of media, so he renamed what he first proposed as the *photophone* the *radiophone* and coined *thermophone* and *actinophone* to describe devices using thermal and actinic rays (Bell 1881, 32, 37). The translatability of sound into electricity, heat, light, and other parts of the electromagnetic spectrum laid bare the principle of analog media, or the significant feature that media could reproduce formal similarity from one medium to another. It was this feature, which, passing through the development of sound on film, would come to form the basis of early electromechanical information systems.

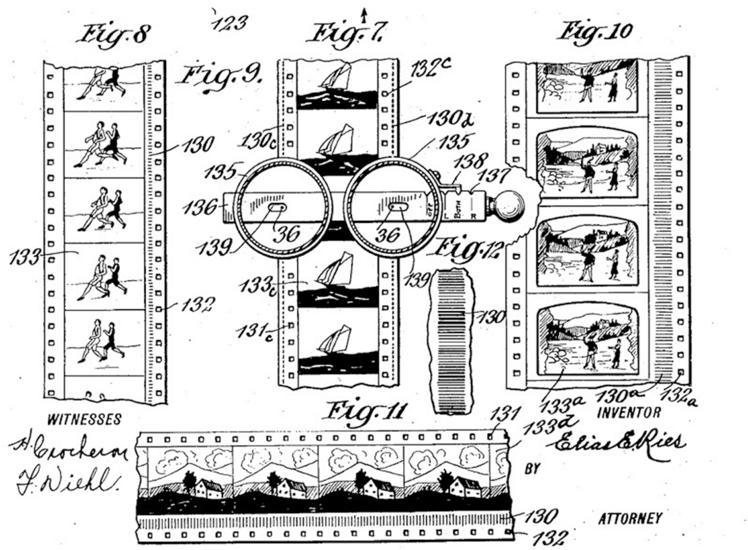
In 1888, Eugene Lauste conceived of a prototype for the optical film soundtrack after reading an article about Bell's photophone. Lauste had been trained in Thomas Edison's laboratory and at the time worked under Edison's lead motion picture engineer, W. K. L. Dickson. First he had the idea to use a band of bromide paper to register sound photographically, but after seeing George Eastman's new commercially available film, which was developed in the same year, he made the switch to celluloid around the time that Dickson was phasing out paper strips in favor of using film for motion pictures. In this way, sound and image alighted on celluloid at the very same moment, even if sound would have to wait another thirty years for a commercially viable system. In 1910, Lauste created the first successful experiment with photographed sound, using a variable density method, a specimen of which was later included in the museum of the Bell Telephone Laboratories (Crawford 1931, 636).

At a time when Edison was decades into his attempt to synchronize moving pictures with the phonograph, as shown in his sound test *Nursery Favorites* (1913), E. E. Ries ([1913] 1926) submitted the first patent application for a variable density method of recording sound on film (fig. 1).[1] Lauste had used a variable area

1 The advantage of sound-on-film over sound-on-disc synchronization is that it virtually guarantees proper synchronization by combining sound and image records on the same medium.

method, in which a "light valve" modified the area where the film was exposed and left a waveform pattern on the soundtrack. The variable density method followed from the experiments of Bell and his associate Charles Sumner Tainter. Instead of varying the area where the light hit the film, it varied the intensity of light, leaving bands of varying brightness on the soundtrack. Significantly, both methods could be reproduced by the same equipment using the same photocells, so that Lauste's method was not entirely lost, returning in the late 1920s as the standard for RCA Photophone in its competition with Fox-Case Movietone's variable density method.[2]

Figure 1: Ries Sound on Film System



Source: Elias E. Ries, US Patent 1,607,480, filed May 21, 1913, and issued November 16, 1926.

That Edison was still trying to synchronize film and sound records on separate media shows a certain poverty of insight, surely due to his own investment in the phonograph.

2 A demonstration of the two recording methods and how they can be run on the same equipment appears in the 1943 ERPI Classroom Film, *Sound Recording and Reproduction (Sound on Film): An Instructional Sound Film*. Along with the Max Fleischer cartoon *Finding His Voice* (1929) and the Vitaphone demonstration *The Voice from the Screen* (1926), this film makes up part of the industry's important effort to naturalize the technology of sound-on-film by explicating its function.

In the annals of sound-on-film, Ries has received little acclaim for his system because, like Lauste before him, he was stuck with the technological dead-end of the selenium cell. He was a lone inventor working on a project that required, when it was finally accomplished, all the might of the professional research laboratories. The complexity of the problem appears in his patent application for "reproducing photographic sound records":

> To reproduce such a record, I employ a method (which is the subject matter of the present application) in which light rays of constant luminosity are projected through an apertured screen similar to the screen employed in making the record, and the record film is moved constantly at a uniform speed in such relation to the aperture, that only an area equal to the area of the aperture will be exposed to the light rays, and the light rays passing through the record film of varying opacity will be projected upon a light sensitive cell or plate, such as selenium. This cell is connected in an electric circuit with a sound reproducing device or telephone, and in accordance with the variations in light rays passing through the record, the light sensitive cell will produce variations in the resistance or cause varying impulses in said circuit to actuate said sound reproducing device or telephone. (Ries 1926, 4)

Ries's sound-on-film system made several important metaphorical leaps: that sound could be transformed into electricity and back into sound (the underlying principle of the telephone); that electricity could be transformed into light and back into an electrical current (the principles of the lightbulb and the selenium cell); and that light could create a record on film capable of being read again by light (the principle of motion pictures). It was an ingenious assembly of metaphors. The trouble was that each metaphor required significant improvements that no single inventor in Ries's time could make alone. As long as he was using a telephone for recording and playback, Ries would never be able to achieve sufficient frequency range and amplification, while the telephone industry had little reason to modify its own technology since it was more than adequate for transmitting the voice intelligibly. Radio and public address would provide early improvements to microphones and loudspeakers in the 1920s, but by the 1930s the film industry would have to develop its own sound technologies, such as Harry F. Olson's invention of the cardioid directional microphone, the shotgun mic, and improved loudspeakers. Having access to only standard light bulbs, Reis would never achieve sufficient luminosity, later supplied by Western Electric bulbs designed specifically for sound-on-film. In 1922, after Lee de Forest bought Ries's patent, de Forest would have to consult with Eastman Kodak about producing a finer-grained film stock to eliminate noise on the soundtrack, an effort that would go on for decades. First and foremost, Ries needed better photocells, the kind of photocells that were going to be developed specifically for reading optical records on film.

In 1918, Lee de Forest began work on what would become Phonofilm, the first commercial application of Ries's patent. Early on, he encountered many of the same limitations that Ries had. After a year of unsuccessful and, it should be noted once again, dangerous efforts given the flammability of nitrate film, de Forest discontinued his "speaking flame" and modified his system to use electrical components, replacing open flames with "light tubes," adding his Audion tube, and testing different photocells (Adams 2012, 219–229). The following year, de Forest contacted Theodore Case about a recent article the latter had published describing the Thalofide cell, a new photoelectric cell growing out of Case's work for the Navy in World War I. Beginning with their correspondence in 1917, de Forest and Case entered a period of professional cooperation, where de Forest brought an overall picture of the sound-on-film idea and Case provided de Forest with invaluable electrical components. Case had begun experimenting with a selenium cell as early as 1911 while still a student at Yale, "with the idea in mind of photographing sound waves," as he relates in a letter to his mother (quoted in Sponable 1947, 284). But he quickly turned away from selenium, adopting a combination of thallium, oxygen, and sulfur that, when used with a vacuum tube, provided an advantageous gain in recovery time, which was "extremely fast," as Case notes in his article, "in marked contrast to selenium" (1920, 290; see also Gomery 2005, 47). Case was not entirely satisfied with his photocell, however, as he notes in a letter to de Forest: "The worst drawback of the Thalofide cell is the hissing noise when exposed to too much light" (as quoted in Adams 2012, 242). And toward the end of his relationship with de Forest, Case improved the device by using potassium. Two *New York Times* headlines registered the significance of Phonofilm in touting its seamless analogical transfer of sound into light: "New Talking Film Photographs Voice" (August 29, 1922); "He 'Photographs' Sounds" (August 17, 1922). But when de Forest failed to mention Case's contributions to Phonofilm in the publicity for its premiere, a dispute arose that led to the dissolution of their relationship. The fruits of their collaboration, however, would continue to grow, and in unexpected ways.

The commercially successful sound-on-film system that Case later developed with Earl Sponable out of this research, the Fox-Case Movietone system, would compete with RCA Photophone and its variable-area method before giving way to Western Electric's optical sound system in 1931. Over the years from de Forest's Phonofilm to the mature sound systems of the 1930s, improvements were made both to photosensors and to film, with simple photocells like the Thalofide cell succumbing to more advanced photomultipliers, such as the RCA 935, and with film stock becoming finer grained. These improvements also facilitated a range of new film data practices in analog computing, microfilm search and retrieval, and mechanical control.

The Transcoding of All Media

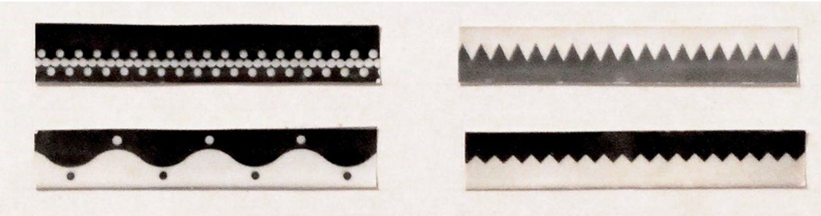
In October 1921, de Forest left his home on the Hudson, in New York State, for Berlin where he hoped to get away from the distractions of his ongoing patent disputes, and perhaps to continue his sound-on-film experiments without tipping his hand to other inventors (see Adams 2012, 238–241). Around the same time in Berlin, Hans Richter was working on his seminal abstract film *Rhythmus 21*. It was Richter's first foray into cinema. Spurred on by his collaboration with Viking Eggeling, and inspired by Eggeling's visionary approach to geometrical form, Richter sought to develop a means of universal visual expression. He had been working since the late 1910s with color dynamics and attempts at visual rhythm, carrying on the tradition of color organs and painted music, which Fred Collopy observes in the titles of Richter's paintings from this period: "*Cello*, *Prelude*, *Fugue*, *Rhythmus 23*, and *Orchestration of Color*" (Collopy 2000, 359). *Rhythmus 21* continued these experiments in visual music, now with the addition of motion. As with de Forest, Richter was making sound with light. The difference was that Richter assumed no need to transduce sound into light. The two were instead direct expressions of a more basic "universal form perception." What mattered most of all was the pattern, something that could inhabit any sense domain, translating between sound, light, and movement, as though music rang out through all the spheres.

Richter's works, however, tested only one side of this analogy between light and sound, showing how rhythm and musical form could be expressed in images. The German-born abstract animator and filmmaker Oskar Fischinger, on the other hand, created what Richter had only theorized about. Fischinger's "sounding ornaments," as well as Rudolf Pfenninger's similar synthetic sound experiments (see Levin 2003), used the optical soundtrack to translate images into sounds (fig. 2). Together with his films *Spirals* (1926) and *Allegretto* (1936–1943), which illustrate the idea of sound rhythms transformed into light, Fischinger's ornament sounds give practical application to the notion that form could be universal across media. Writing in 1932, Fischinger explains the "purity" of these "sounding ornaments":

> Between ornament and music persist direct connections, which means that Ornaments are Music. If you look at a strip of film from my experiments with synthetic sound, you will see along one edge a thin stripe of jagged ornamental patterns. These ornaments are drawn music – they are sound: when run through a projector, these graphic sounds broadcast tones or a hitherto unheard of purity, and thus, quite obviously, fantastic possibilities open up for the composition of music in the future. (Fischinger 1932, n.p.)

In a sense, Fischinger took de Forest's technology and cut it in half, doing away with the messy business of capturing images and sounds from "out there" in the world, and simply generating them ex nihilo, in a way that created sounds that never could have existed otherwise. If an apparatus could transform light into electricity into sound based on the differential density on a film soundtrack, it could transform any visual shape or density so inscribed on the track. Fischinger was, so to speak, taking what had been a problem for sound designers in Hollywood – the fact that the apparatus would render any "noise" inscribed on the soundtrack, whether scratches or specks or graininess of the film – and turning it into a new form of artistic expression.

Figure 2: Fischinger Drawings



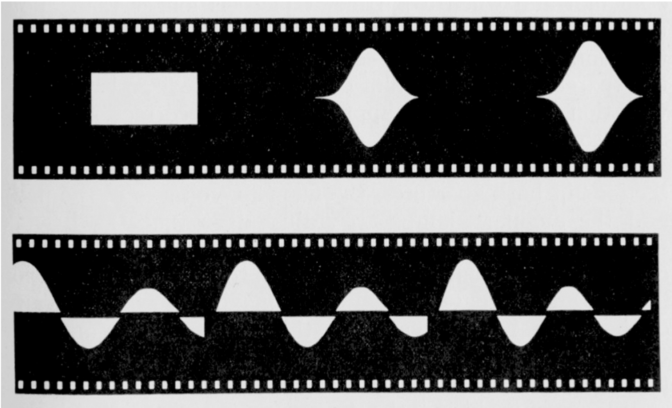
Source: Collection and © Center for Visual Music, Los Angeles.

What is striking about Fischinger's sounding ornaments is that some of the strips are nearly indistinguishable from the control patterns used in an early analog computer, the Cinema Integraph, created by Gordon S. Brown at MIT (fig. 3). The idea of using motion picture film to perform computations more complex than those calculated by Vannevar Bush's Differential Analyzer is said to have occurred to Norbert Wiener while he was enjoying a night at the theater:

> It was the old Copley Theater in Boston, and I'd been thinking very much. You see, I'd been tremendously inspired by Vannevar Bush's work on his various sorts of computing machines, and I thought I'd get a hit in for my own. The optical machine was conceived during the intermission there, and it was taken up and pushed by Bush and by various people whom he lent me, including Dr. Brown. (Brown and Wiener [1955] 1984, 379)

The Cinema Integraph was designed to perform harmonic analysis, which involved more complicated functions than the differential equations performed by Bush's mechanical analyzer. As Wiener summarizes, the problem came in having "numbers distributed over a plane, or over a volume." "If you wanted to solve a problem whose answer was distributed in more dimensions," he recalls, "you had to be able to represent a function in more dimensions" (Brown and Wiener 1955, 380). Wiener's insight was to turn to "the density of a photographic negative, which varies up and down and left and right," utilizing both the vertical and horizontal dimensions of film as variables and in this way capitalizing on the mechanical movement of the reel as well as the photographic space of the frame (Brown and Wiener 1955, 380). As with Fischinger's ornament sounds, the patterns of the Cinema Integraph could be read by an optical sound reader to create noises. It should be no surprise, then, that the optical soundtrack was already being used as a data medium in the 1930s.

Figure 3: Cinema Intergraph



Source: H. L. Hazen and G. S. Brown, "The Cinema Integraph: A Machine for Evaluating a Parametric Product Integral," *Journal of the Franklin Institute* 230, no. 1 (July 1940), p. 35.

Richard S. Morse, who began his career at Eastman Kodak in 1935 before going on to achieve his greatest fame for being the inventor of frozen orange juice concentrate and founder of the Minute Maid Corporation, sought to bring order to the noise of microfilm's accumulating data trail. In a sense, Morse was doing what Fischinger was doing, creating music out of noise, trying to organize the soundtrack of the modern world. Morse made several contributions to sound-on-film technologies while at Kodak, including an amplification system for home movies, "using a radio receiver to amplify the output of the photoelectric cell of a sound motion picture projector" (Morse 1938, 3), as well as a system for high fidelity recording that used compression during recording and expansion during replay, much the same as an anamorphic lens compresses and expands widescreen images (Morse 1939). He also established important advances in push-pull recording, which used multiple soundtracks side-by-side along the edge of the film to reduce background noise and allow for greater amplification (McLeod and Morse, 1939; see also Hilliard 1938 and Ceccarini 1938). The patent he filed in 1938 for a "Rapid Selector-Calculator" developed out of this work. Its variable density coding system inspired by optical motion picture soundtracks came to be called, in the small circle of microfilm search and retrieval, a data soundtrack (Morse 1942, 1).[3]

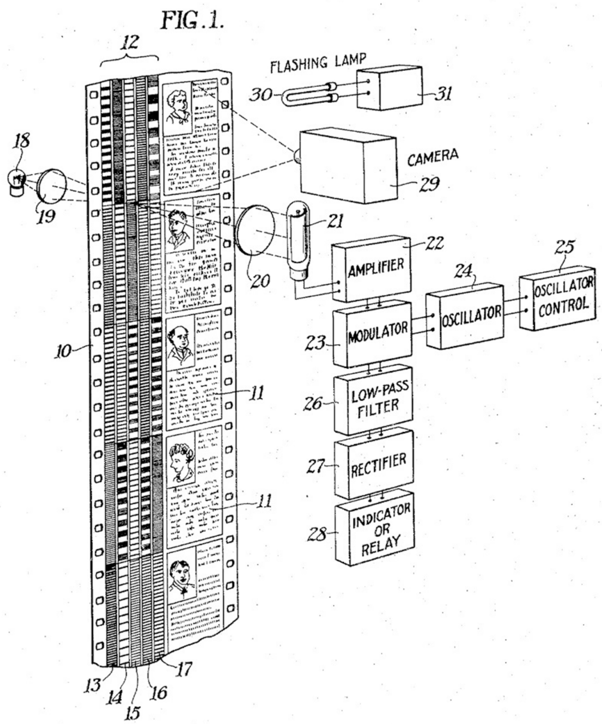
Morse's data soundtracks (fig. 4) promised the automatic sorting, selecting, and calculating of records on microfilm so that records could be found even if the operator did not know their place on the film. The system specialized, in Morse's account, in items such as literature, bank checks, sales records, and personal identification. The device used high-speed photography to capture a duplicate of an "information item" – a single frame on the image track – which made it possible to retrieve multiple items very quickly. Morse gives one example of the machine's "almost innumerable" applications:

> A bureau of criminal investigation is provided with a band carrying either five tracks corresponding to the file numbers of registered criminals or a photographable area showing the number and name of the criminal. Adjacent to these tracks or area is a plurality of characteristics of the criminal such as city, state, or area of activities, type of activity, color of hair, height, approximate year of birth, etc. When a description of a criminal and/or a crime is reported, and corresponding decoding devices (such as oscillators) arranged, the band provides a list of likely suspects in an automatic manner. (Morse 1942, 4)

As Fischinger had done in capturing the fugitive noises of the soundtrack, Morse imagined that data soundtracks captured criminal attributes in an information net. But the device's stated applications made it not simply a benign labor-saving technique. As Morse's description shows, the use of film in data applications sought to expand the power of the carceral state, prefiguring the techniques of predictive policing and racial identification that would characterize the data practices of the later computer revolution (see McIlwain 2019; Benjamin 2019).

3 Morse mentions that the data track might be better termed a "frequency track," but given the analogy with sound-on-film systems the notion of a data soundtrack continued to be the common descriptor in subsequent literature. See, e.g., Burke 1992, 652.

Figure 4: Morse Data Soundtracks



Source: Richard S. Morse, "Rapid Selector-Calculator," US Patent 2,295,000, filed June 23, 1938, and issued September 8, 1942.

Morse's optical technology in a sense brought Hollerith's "punch photograph" full circle once again: the punched card that simulated the optical medium of the photograph returned to the optical medium of film in order to retrieve the data of the punched card. In fact, the punch system of the Hollerith card had already been used in an earlier microfilm search and retrieval system. In 1931, Emanuel Goldberg, inventor of the Kinamo portable camera that Joris Ivens used to film *The Bridge* (1928), presented his Statistical Machine at the Eighth International Congress of Photography in Dresden. The information retrieval system connected with Goldberg's presentation from the night before, when he detailed a new sound-on-film technology developed at Zeiss Ikon. The Statistical Machine was, in biographer Michael Buckland's words, "a revolutionary document search and display system using microfilm for document storage, a photoelectric cell for sensing index codes, and digital circuits for pattern recognition" (Buckland 2006, 155).[4] It was designed to retrieve sales data for business, and used a metadata system very similar to what Morse later used in his Rapid Selector-Calculator.[5] Each record contained a code with certain characteristics, just like Hollerith's punch photographs only now automated on a soundtrack. Goldberg and Morse thus solidified Wiener's intuition that optical computing could control automation, that cinematics could control kinematics (see Stine 2014). Before McLaren's animated sound strips, data soundtracks had already taken up the cybernetic task of animating machines.

At the basis of the experiments of de Forest and Case, Eggeling and Richter, Morse and Goldberg, Fischinger and Pfenninger, Wiener and Brown, is a world conceived in advance as calculable and capable of being translated from one medium to another. That physical materials or forms of energy were expressible in terms of each other was only one small part of this understanding. If the world was conceived in advance as calculable, then all these physical materials and forms of energy – sound, light, electricity, machine movements, etc. – were expressible not only in terms of each other but also universally expressible, by abstraction. Eggeling and Richter understood this from the beginning. Visual music was only one small part of the "universal form expression" they sought to capture through their artistic work. In 1920, the two collaborated on a pamphlet advocating the development of a universal form language "above and beyond all national language frontiers" (as cited in Turvey 2003, 26).[6] The pamphlet expressed a desire that Eggeling had been carrying with him most of his life, and which he promoted in Richter, of creating "a new communication machine" (O'Konor 1971, 7). It would not be long before science would make their dream of a universal language a genuine technical possibility. In 1937, shortly after Fischinger's sounding-ornaments experiments, Alan Turing settled once and for all that these different forms of energy were expressible in a common language with the Universal Turing Machine. The Turing concept would soon make it possible for a single machine to store, process, and transmit all media, whether sound, image, or text.

4 Buckland deserves recognition for restoring Goldberg's place in the history of information science, which was nearly forgotten in the shadow of Vannevar Bush's Memex. The biography also outlines Goldberg's important contributions to photography, cinema, and sound-on-film.

5 Other applications Buckland highlights are check handling and the preparation of telephone subscriber invoices (Buckland 2006, 148).

6 From Hans Richter's retrospective remarks on their collaboration. As Turvey notes, the original text of *Universelle Sprache* ("Universal Language") is probably no longer extant.

A Control Medium for Machine Automation

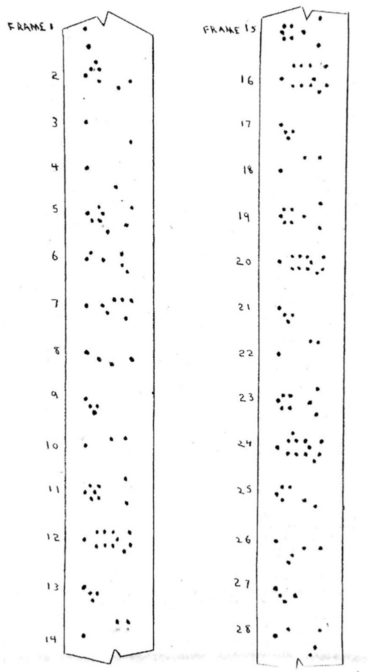
When engineers and inventors set out to develop automatic machine tooling, they turned to the same readily available assemblies of photocells and motion picture film as information scientists had. In his description of the problems facing machine tooling in the 1950s, Frederick W. Cunningham explains: "All of these machines have the disadvantage of requiring that a master piece or, in the automatic screw machine, a set of master cams, be made and installed every time the machine is to be changed from one job to another" (1954, 487). Because of the wear inflicted by the work process, master templates had a short life span, as David F. Noble notes: "Storage of templates, most of which were for a single job or a single part of a job, was costly and required complex inventory and retrieval systems" (1986, 83). To Cunningham's mind, it was necessary to find a control medium for shaping metal parts that would not itself be damaged in the work process. Out the possibilities of magnetic tape, punched cards, and film, he selected film because of its durability, widespread availability, and cost effectiveness.

In a sense, motion pictures had supplied the means of automating interchangeable programs from the beginning. A film projector is capable of showing any film, or in a sense running any program (it is worth recalling that evenings at the movies were historically called "programs"). This ability also made film an especially good medium for controlling machine movements because it was itself motion-controlled by the sprocket holes lining its sides and was intentionally durable for such a purpose. Bazin was right to point out these automatic features of the technology: "For the first time, between the originating object and its reproduction there intervenes only the instrumentality of a nonliving agent" ([1945] 1967, 13). Cinema was early automation. The programs it ran could be projected for people or as clandestine matinees intended only for machines. In this light, Noble's (1995) notion of "progress without people" might be refigured as "cinema without people" because the same technologies that have entertained mass audiences have been just as easily configured to run machines.

Across several fields, researchers realized this advantage of using film to change the program of automatically controlled machines. Cunningham selected 16mm film to solve the problem of cutting non-circular gears, such as those needed for the Army T-41 Rangefinder anti-aircraft gun. Non-circular gears were difficult to produce, even for a skilled machinist. The problem became so great in the aerospace industry that by the end of the 1950s, noncircular and non-rectangular surfaces necessitated automation both in manufacturing and design, a difficulty that would form the thrust of early computer-aided design systems. Cunningham's solution involved an assembly of electronics and photocells set up to read a complicated machining program printed on film, "a continuous stream of data that would dictate the immediate movement of machine elements" (Ashburn 1953, 150). Several other experiments followed similar pursuits. The Swiss firm Contraves A.G. improved on a German machine developed during the war to produce a photo-optical tracer similar to that devised at MIT (see Noble 1995, 83; Hazen, Brown, and Jaeger 1936). Cletus Killian, drawing on his work in developing mechanical computing machines at Remington Rand, similarly experimented with a photo-optical line-following machine, which he called the "Automatic Machinist" (1952, 1). Killian's patent for the machine explicitly lists cellulose acetate and cellulose nitrate film as excellent control tape (1952, 14).[7] Alongside its uses as an analog photographic medium, film also served as punched tape for running machines digitally. Albert Gallatin Thomas, an MIT alumnus who worked with Vannevar Bush in the 1920s, followed Killian's approach to produce a discretely coded system capable of converting digital information, stored on film, into continuous three-axis machine movement (Noble 1986, 87). Extending automatic machining capabilities, F. P. Caruthers developed a system at Thomson Equipment that employed motion picture film as punched tape (fig. 5) in conjunction with a plugboard to control four axes of machine motion (Noble 1986, 92–96; Carr 2014, 271–273). Noble argues in support of Caruthers's method that, unlike MIT's numerical approach, it was designed with machinists in mind, not to replace them. Machining with the system remained a sensory experience rather than an attempt to circumvent the senses: "Through the use of dials which permitted both coarse and fine tuning, the operator could set and adjust feeds and speeds, relying upon accumulated experience with the sights, sounds, and smells of metal cutting" (Noble 1986, 94).

7 The application substituted for an earlier abandoned application dating from May 18, 1943. As Noble notes, however, Killian's work was largely a failure (1986, 87).

Figure 5: Caruthers Punched Film



Source: Caruthers, Felix Porter. 1984. *Automatic Machine Control: A Five Generation History of Numerical Control Systems as Conceived and Developed by Felix Porter Caruthers*. Three Rivers, CA: Caruthers & Associates, Inc., 1984. Smithsonian Institution Archives Center, National Museum of American History, Felix P. Caruthers Papers, 1952–1991, Box 1, Folder 1, n.p.

At the same time, the driving force in photocell innovation was the film industry, where "the principal contemplated application was sound-on-film pickup" (Engstrom 1980, 3–4). Audiences who paid for tickets to the talkies, unwittingly perhaps, drove the demand for photocell applications in a variety of other fields, including machine automation. Although military exploits have often been cited as the very visible hand behind technological growth, they played at best a tangential part in this unfolding where ideological productions screened from view a more pervasive technological pursuit. By the 1940s, photocells were being used for process control in industry, as sensors for the automatic doors at Penn Station in New York, and in two feedback-controlled automatons, Henry Singleton's Moth and Bedbug and Grey Walter's tortoises. As ought to be expected with such feedback systems, it was only a matter of time before developments returned to the entertainment industry with the "total automation of the motion-picture theater" (Boudouris, Gray, and Burlinson 1972, 81).

Conclusion

As this history suggests, the feedback loop between today's cinematic practices and Big Data, such as we see in algorithmic video platforms like Netflix and YouTube, was already in the works in the classical era of film, even if the circuit of their interchange was long and generally imperceptible. Visual media scaled between the two-dimensional world of linear code and the three-dimensional world of machine action, enabling new data practices and the mobilization of these data into new mechanical systems. Film was used to reproduce, most obviously, images, but it also became the support through these images to reproduce any number of other media, such as sound and text. Specifically in terms of its reproduction and search and retrieval of text, film in this way became practical in libraries, hospitals, and police offices, as a means of information processing before digital computing. As Lev Manovich notes, film achieved its greatest effect in computing by obliterating the image, using film not as a pictorial space, but in the case of Konrad Zuse's Z3 computer disregarding the image in the form of a punched-hole code (2002, 20). But as Kittler notes, the binary alternation effectuated by punched holes was part of the medium of film from the beginning in its use of negative–positive photography:

> The consequences of unlimited copying are clear: in a series first of originals, second of negatives, and third of negatives of a negative, photography became a mass medium. For Hegel, the negation of a negation was supposed to be anything but a return to the first position, but mass media are based precisely on this oscillation, as it logically calculated Boolean circuit algebra and made possible nothing less than the computer. (2010, 134)

So it should come as no surprise that the functions of computing were first tested on the medium of film. From Norbert Wiener and Gordon S. Brown's Cinema Integraph to Richard S. Morse's data soundtracks and Emanuel Goldberg's Statistical Machine, in Fredrick Cunningham's noncircular gear shaper and F. P. Caruthers's punched-film machinist, motion picture film was a computing medium that prefigured much of what would be taken up by digital computers. It was a universal medium operating in an era on the cusp of universal computing machines.

References

Adams, Mike. 2012. *Lee de Forest: King of Radio, Television, and Film*. New York: Springer Science+Business Media.

Ashburn, Anderson. 1953. "Film Runs Non-Circular-Gear Shaper." *American Machinist*. February 2, 1953: 149–150.

Bazin, André. (1945) 1967. "The Ontology of the Photographic Image." In *What Is Cinema?* Vol. 1. Translated by Hugh Gray, Berkeley: University of California Press, 9–16.

Bell, Alexander Graham. 1880. "The Photophone," *Science* (1): 132.

———. 1881. *Upon the Production of Sound by Radiant Energy*. Washington, DC: Gibson Bros.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity.

Boudouris, Al, Geoffrey T. Gray, and John J. Burlinson. 1972. "Total Automation for the Motion-Picture Theater," *Journal of the Society of Motion Picture and Television Engineers* 81 (2): 81–87.

Brown, Gordon S., and Norbert Wiener. (1955) 1984. "Automation, 1955: A Retrospective." *Annals of the History of Computing* 6 (4): 372–383.

Buckland, Michael. 2006. *Emanuel Goldberg and His Knowledge Machine: Information, Invention, and Political Forces*. Westport, CT: Libraries Unlimited.

Burke, Colin. 1992. "The Other Memex: The Tangled Career of Vannevar Bush's Information Machine, the Rapid Selector." *Journal of the American Society for Information Science* 43 (10): 648–657.

Carr, Nicholas. 2014. *The Glass Cage: Automation and Us*. New York: Norton.

Case, T. W. 1920. "'Thalofide Cell' – A New Photo-Electric Substance." *Physics Review* 15 (4): 289–292.

Ceccarini, O. O. 1938. "Theoretical Notes on the Push-Pull Method of Recording Sound." *Journal of the Society of Motion Picture Engineers* 30 (2): 162–168.

Chun, Wendy Hui Kyong. 2008. "The Enduring Ephemeral, or the Future Is a Memory," *Critical Inquiry* 35 (1): 148–171.

Collopy, Fred. 2000. "Color, Form, and Motion: Dimensions of a Musical Art of Light." *Leonardo* 33 (5): 355–360.

Crawford, Merritt. 1931. "Pioneer Experiments of Eugene Lauste in Recording Sound," *Journal of the Society of Motion Picture Engineers* 17 (4): 632–644.

Cunningham, Frederick W. 1954. "Controlling Machine Tools Auto-matically." Mechanical Engineering 76 (6): 487–490.

Engstrom, R. W. 1980. *RCA Photomultiplier Handbook*. Lancaster, PA: RCA Corporation.

Faraday, Michael. 1991. *The Correspondence of Michael Faraday: 1841–1848*, vol. 3, edited by Frank A. J. L. James. London: The Institute of Engineering and Technology.

Fischinger, Oskar. 1932. "Sounding Ornaments." *Center for Visual Music*. Accessed September 16, 2021. http://www.centerforvisualmusic.org/Fischinger/SoundOrnaments.htm.

Gaboury, Jacob. 2021. *Image Objects: An Archaeology of Computer Graphics*. Cambridge, MA: MIT Press.

Gitelman, Lisa. 2014. *Paper Knowledge: Toward a Media History of Documents*. Durham, NC: Duke University Press.

Gomery, Douglas. 2005. *The Coming of Sound: A History*. New York: Routledge.

Hazen, H. L., and G. S. Brown. 1940. "The Cinema Integraph: A Machine for Evaluating a Parametric Product Integral," *Journal of the Franklin Institute* 230 (1): 19–44.

"He 'Photographs' Sounds." 1922. *The New York Times*. August 17, 1922: 6.

Hilliard, J. K. 1938. "Push-Pull Recording." *Journal of the Society of Motion Picture Engineers* 30 (2): 156–161.

Hollerith, Virginia and Herman Hollerith. 1971. "Biographical Sketch of Herman Hollerith." *Isis* 62 (1): 69–78.

Killian, Cletus H. 1960. "Automatic Machinist." US Patent 2,947,928, filed May 18, 1943, and issued August 2, 1960.

Kirshenbaum, Matthew. 2008. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: MIT Press.

Kittler, Friedrich. 2009. "Towards an Ontology of Media." *Theory, Culture & Society* 26 (2–3): 23–31.

———. 2010. *Optical Media: Berlin Lectures, 1999*. Translated by Anthony Enns. Malden, MA: Polity.

Levin, Thomas Y. 2003. "'Tones from out of Nowhere': Rudolph Pfenninger and the Archaeology of Synthetic Sound." *Grey Room* 12 (12): 32–79.

Manovich, Lev. 2002. *The Language of New Media*. Cambridge, MA: MIT Press.

Charlton D. McIlwain. 2019. *Black Software: The Internet and Racial Justice, from the AfroNet to Black Lives Matter*. Oxford: Oxford University Press.

McLeod, John H. and Richard S. Morse. 1939. "Sound Recording." US Patent 2,161,369, filed October 13, 1937, and issued June 6, 1939.

Morse, Richard S. 1938. "Sound Reproducing Apparatus." US Patent 2,129,227, filed October 14, 1936, and issued September 6, 1938.

———. 1939. "Sound Motion Picture Reproducing Apparatus." US Patent 2,180,014, filed October 23, 1936, and issued November 14, 1939.

———. 1939. "Sound Reproduction." US Patent 2,158,193, filed September 30, 1937, and issued May 16, 1939.

———. 1942. "Rapid Selector-Calculator." US Patent 2,295,000, filed June 23, 1938, and issued September 8, 1942.

Müller-Sievers, Helmut. 2011. *The Cylinder: Kinematics of the Nineteenth Century*. Berkeley: University of California Press.

"New Talking Film Photographs Voice." 1922. *The New York Times*. August 29, 1922: 1.
Noble, David F. 1986. *Forces of Production: A Social History of Industrial Automation*. New York: Oxford University Press.
———. 1995. *Progress without People: New Technology, Unemployment, and the Message of Resistance*. Toronto, ON: Between the Lines.
O'Konor, Louise. 1971. *Viking Eggeling, 1880–1925: Artist and Filmmaker, Life and Work*. Translated by Catherine G. Sundström and Anne Bibby. Stockholm: Almqvist & Wiksell.
Otlet, Paul. (1925) 1990. "The Preservation and International Diffusion of Thought: The Microphotographic Book." In *Selected Essays of Paul Otlet*, translated and edited by W. Boyd Rayward, 204–210. Amsterdam: Elsevier.
———. (1901) 1990. "On a New Form of the Book: The Microphotographic Book," in *Selected Essays of Paul Otlet*, translated and edited by W. Boyd Rayward, 87–95. Amsterdam: Elsevier.
Ries, Elias E. 1926. US Patent 1,607,480, filed May 21, 1913, and issued November 16, 1926.
Rossell, Deac. 1998. *Living Pictures: The Origins of the Movies*. Albany: State University of New York Press.
Sponable, Earl I. 1947. "Historical Development of Sound Films." *Journal of the Society of Motion Picture Engineers* 48 (4): 275–303.
Stine, Kyle. 2019. "Critical Hardware: The Circuit of Image and Data." *Critical Inquiry* 45 (3): 762–786.
———. 2014. "The Coupling of Cinematics and Kinematics." *Grey Room* 56 (4): 34–57.
Thylstrup, Nanna Bonde. 2019. *The Politics of Mass Digitization*. Cambridge, MA: MIT Press.
Turing, Alan Mathison. (1947) 2004. "Lecture on the Automatic Computing Engine." In *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus the Secrets of Enigma*, edited by B. Jack Copeland, 378–394. Oxford, UK: Oxford University Press.
Turvey, Malcolm. 2003. "Dada Between Heaven and Hell: Abstraction and Universal Language in the *Rhythm* Films of Hans Richter." *October* 105 (3): 13–36.
von Neumann, John. (1945) 1993. "First Draft of a Report on EDVAC." *IEEE Annals of the History of Computing* 15 (4): 27–75.

Film Box Office Charts and the Metadata of Culture

Liam Cole Young

Counting culture is now commonplace. Areas of cultural and artistic production once thought distinct from numbers have in recent decades been reconfigured by myriad practices of quantification, data capture and analysis. Sales and streaming charts not only measure the circulation and reach of film, television, music, and podcast texts, but often rank their relative *impact* or *importance*. Meanwhile, algorithmically generated recommendation systems curate individual experiences of streaming and search platforms such that no two users' experiences are the same. In sports, new forms of data capture and analysis track athlete performance, predict game and event outcomes, enable new modes of corporate decision making, and intersect in powerful ways with gambling and financial markets. Closer to home, citation counts and *impact factors* create new incentive structures and transform standards by which academic research is judged relevant, successful, or even legitimate. Such data practices transform traditionally qualitative realms into an unending series of quantifiable units and forms – so many numbers, charts, lists, and spreadsheets to be parsed by ever more "innovative" metrics and analytical systems.

These changes seem novel, part of the wide-ranging consequences of ubiquitous computing and digitization, but in fact they have deeper genealogies. The application of large-scale quantitative methods to measuring and analyzing culture did not begin with the internet, even if Silicon Valley's championing of these methods has accelerated and intensified their normalization within discourses of innovation, growth, and disruption. Any question of digital culture's bias toward analytics must, in my view, situate its privileged techniques of quantification within a longer arc of modernity. Epistemologies associated with Western modernity have counted, compressed, classified, calculated, and circulated culture and much else for a lot longer than the internet has been around. These deeper histories remind us that cycles of struggle have long unfolded at the intersections of media-technologies and class, racialization, colonialism, and epistemology. Recent scholarship has focused on, for instance, cultural, as opposed to state, modes of counting through such topics as pop music charts (Huber 2010; Straw 2015), timing and imaging in sport (Finn 2016) credit bureaus and reporting (Lauer 2017), media and multiculturalism (Hayward 2019), files, filing cabinets, and filing hands (Vismann 2008; Robertson 2021), feminist histories of quantification (Wernimont 2019), payment systems (Swartz 2020), and "methodolatry" (Mattern 2013). Earlier examples include texts on charts by Parker (1991) and Hakanen (1998), on memos by Yates (1993) and Guillory (2004), or infrastructures of classification such as covered by Bowker and Star (2000). In spite of this vibrant interest in quantitative cultures and institutions, there is surprisingly little research on how such techniques structure activity within cultural industries.

This essay contributes to this growing body of critical literature by tracing the emergence and rise of Hollywood box office charts across the North American media landscape of the late 20th century. Over a roughly fifty-year period, from the late 1960s to today, box office charts and other informational forms, such as pop music charts, professional sports tables, and University rankings have played a major role in reconfiguring the epistemological ground of culture. These forms materialize certain *data practices* – modes of observing, measuring, visualizing, comparing, and ranking – and are an important part of the story of how realms historically thought separate, or at least insulated, from numbers and market-based pressures became subject to prevailing modes of quantification and financialization.

During this period, box office charts transformed from an afterthought on Hollywood's margins into a dominant industrial and institutional form.[1] Today, such charts are perhaps the single most important metric by which film performance – whether of individual films, actors, filmmakers, or the industry as a whole – are discussed and assessed, from within Hollywood and without. Charts embody an ecosystem of tracking and analysis that is used not only in reactive ways, such as in decision-making contexts, but also for prediction and speculation. The small academic literature on the paper-based charts discussed in this essay does not typically question the nature of the informational form itself, i.e., how and why charts assemble the data they do in the ways they do. The typical view of paper-based box office charts accepts charts as "a concise and condensed representation of the audience's behavior patterns" (Buckland and Long in Mathijs 2006, 92) and thus rests on an assumption that such forms simply gather and reflect "informa-

1 The most comprehensive account of this transformation, to my knowledge, is Hayes and Bing's 2004 book, *Open Wide: How Hollywood Box Office Became a National Obsession*. Though a work of journalism that focuses on the *fin de siècle corporate* developments in Hollywood related to opening weekends, the authors are unusually attentive to systemic forces and historical processes that anticipate contemporary developments. "Box office, then, isn't just a story of studios, market research, advertising and movie stars; it is the story of the architectural and retail infrastructure into which the studios plug their product" (Hayes and Bing 2004, 13). I am greatly indebted to their archival and interview-based research.

tion" that exists *a priori* in the world.[2] My goal here, following Latour's "first rule of method," is to disrupt such assumptions by "un-black boxing" box office charts (Latour 1987, 21), considering instead how they function as what John Durham Peters (2012; 2015) calls *logistical media*. The latter are forms and practices that abstract, order, organize, and compute "basic coordinates of time and space," establishing the "central points around which culture rotates" (Peters 2012, 41–42). His examples are calendars, clocks, and towers, but others have pushed this definition to explore such case studies as radar (Case 2013), digital calendars (Wajcman 2019), thumbnail images (Thylstrup and Teilmann-Lock 2017), Amazon's anticipatory shipping patent (Nyckel and Poechhacker 2020), as well as more general categories of software (Rossiter 2016), smart cities (Shapiro 2018), police media (Reeves and Packer 2013), and formats (Volmar 2020). Such studies show that logistical media do not simply record, represent, or reflect culture but actively produce the distinctions, categories, and concepts upon which its epistemological assumptions rest.

Like pop music charts, we refer to box office figures as "charts" even though they are really just lists: a series of aggregated units ranked in descending order of performance, whatever the metric. Lists are powerful logistical media because they exist at the borderline between entropy and order, "etcetera" and "everything-included" (Eco 2009), processing sense from nonsense (Siegert 2015; Serres 2007 [1980]). Who or what draws these distinctions, how, and for what purposes are essential questions for understanding how human beings think and act, how we stitch things together across space and time, especially when considering lists of people, resources, or orders (Werbin 2017; Young 2017). As Jack Goody (1977) first showed, to make a list is to visualize words, things, events, or people into two-dimensional units that can be combined with others, classified, or otherwise manipulated in ways not possible in thought and oral speech. That box office charts are called "charts," rather than lists, highlights this feature. Disaggregated units, films, are drawn together to form a document that allows people to do things; namely, to chart a course through increasingly noisy information environments. To conceive of charts as cartographic aids, however, is to ignore how they constitute the very ontology of a cultural field that they claim simply to represent or report. In the assembly of charts, data points are generated by practices of observation, measure,

2 Frohmann (2004a; 2004b) argues we should not understand "information" as an ontological substance to be sought, gathered, and translated into meaning (i.e. knowledge) by the human mind. Instead, "information" is a concept that arises only from documentary and data practices that do stabilizing and coordinating work that precedes abstract concepts like "information," but also "objectivity," "fact," and even "knowledge." Documents become informing or informational, he argues, only once they acquire the cache of the "social and pedagogical disciplines that maintai[n] them" (Frohmann 2004a, 400).

counting, and comparison. These are flattened, standardized, aggregated, and organized in such informational forms as box office charts before, finally, being operationalized as knowledge in various ways and by various communities.

Inspired by the praxeological approach of this volume, I seek to *follow the practices themselves*, to emphasize where and how box office data came to be aggregated and displayed, as well as some of the motivations that led people to *datafy* the production, distribution, and reception of cinema in the first place. I hope to show some of the ways such data was put to work, how it came to reconfigure chains of decision making and resource distribution within the film industry. In such uses, charts were pushed beyond functions of measure or comprehension and toward prediction. As such, box office charts offer a crucial step in the genealogy of contemporary cloud-based forms of tracking, prediction, and decision-making as found in recommendation and newsfeed algorithms. This genealogical approach involves illuminating a small segment of a much larger and more complex chain of actors, actions, events, texts, sites, and motivations that produced the data environment typified by box office charts. I thus look upon charts themselves, as crystallizations of particular data practices and imaginaries, but also actors who did the charting, most notably *Variety* critic A.D. "Art" Murphy. He plays a large role in these pages not because I seek to tell an origin story or emphasize Murphy's individual genius or villainy, but because he occupied a privileged site and step in this genealogy. *Variety* played an outsized role in popularizing box office charts, and it was Murphy who developed the magazine's original chart format.

My argument is not that counting culture is in all cases a bad thing, nor that box office charts "caused" the corporatization or financialization of art and culture. Rather, and more humbly, I seek to shine light on the epistemological work performed by informational forms and logistical media, and on the data practices and imaginaries that produce them. Such considerations rarely appear in histories of culture, while accounts of contemporary data analytics typically begin only with the digital computer. My aims are thus historical and deflationary, to show how deeper histories of counting culture prefigure and anticipate today's Big Data, and to weave practices of accounting into stories about cultural industries so as to complement the usual emphasis on production and consumption.

Hollywood Datafication

The now-familiar form of box office charts, a ranked list of the top grossing films over a defined period (usually a week or weekend), took some time to settle into place. The recent proliferation of box office charts online followed a slower but no less robust propagation across so-called "legacy" media environments between about 1970 and 2000. Appearing first in *Variety* in the late 1960s, the charts proved in subsequent decades to be a perfect morsel for content-hungry editors of increasingly specialized entertainment and business presses, whether at longstanding venues like *Variety* and *Forbes*, or newer outlets like *USA Today* (est. 1982), *Premiere* (est. 1987), and *Entertainment Weekly* (est. 1990). Television producers also mined value from charts, particularly those looking to fill airtime for niche entertainment programming that arose with cable and satellite, such as *Entertainment Tonight* (est. 1981), MTV (Music Television, est. 1981) and VH1 (Video Hits 1, est. 1985) (see Hayes and Bing 2004, 294–95). But the earliest box office reporting was much less conspicuous, more infrequent, and lacked any kind of systematic approach. Reports were linked to specific theatres or cities and were not aggregated into larger data sets that might provide insight into, for instance, regional or national trends (Hayes and Bing 2004, 290). These box office reports made few claims to objectivity, such as would later come to be associated with box office charts, and fit neatly alongside the qualitative content with which *Daily Variety* had made its name since launching in 1933[3] such as film reviews, celebrity chatter, and interviews. As the 1960s drew to a close, what little profile box office figures had was in the service of Hollywood gossip rather than economic analysis or prediction.

This would change as Hollywood entered a new epoch characterized by the decline of the studio system that had been relatively stable over the first half of the 20th century. The system, in which five major studios controlled every moment and site along the chain of production, distribution, and exhibition of film commodities, was gradually dislodged by a series of economic, legal, technological, and cultural forces. Some of these were external to the industry, such as the rise of television, and some internal, such as the infamous Paramount anti-trust suit. The latter was settled by the United States Supreme Court in 1948 and required major studios to dismantle their vertically-integrated operations. With studios no longer able to enjoy an oligopoly over talent, financing, and distribution, an array of independent operators elbowed their way into the industry.

I am not treading new ground here. This story is often told by cultural historians and film scholars as context for the 1970s "golden age" of American cinema, during which, to borrow the title of Peter Biskand's (1999) account, a few "Raging Bull" and "Easy Rider" artists, producers, and critics shook Hollywood out of complacency.[4] These young filmmakers reimagined the form, content, and infrastructure of American film. They shot and edited their films differently, adopting

3 New York City-based *Variety* magazine was founded in 1905 as a weekly periodical focusing on vaudeville and theatre. *Daily Variety*, launched in 1933, was a Los Angeles-based subsidiary that focused on the motion picture industry. It ran until 2013.

4 See, for instance, the critical overview comprised by essays in Elsaesser, Horwath, and King (2004).

verité and direct cinema techniques from the French New Wave. They mined diverse source material for stories, exploring the streets and sweat of life in New York City and Detroit as often as high modern literary traditions or low cultural forms such as American vaudeville. By synthesizing arthouse and grindhouse, filmmakers associated with this golden age were beloved by film critics and ticket buyers alike. But it was not to last. The end of the golden age corresponds with the opening of a new period of corporate-commercial "blockbuster" cinema of the 1980s and beyond, after the studios had reclaimed autonomy over the film business. Depending on the storyteller, this transition is framed as one of artistic marginalization and corporate expansion, or as a return of fiscal responsibility in the wake of free spending. What is clear is that the end of the 1970s brought a prolonged and intense period of consolidation, convergence, and financialization (for a comprehensive overview, see deWaard 2020). These processes produced a new series of dominant forms and rituals in the industry, such as the blockbuster or "event" film (Acland 2020); hugely expanded promotional campaigns that included movie tie-in merchandizing, licensing, and soundtrack deals; the sale and renewal of TV broadcast rights; an industry calendar reformatted around various movie "seasons" (summer blockbuster season, awards season, low season, etc.); and, perhaps most important of all, a new emphasis on release dates and opening weekend grosses.

I am less interested in offering a moral argument about whether these changes were good or bad than in highlighting that accounts of this period typically overlook the corresponding changes to techniques of data collection, aggregation, and presentation employed by Hollywood insiders to track and analyze commodity circulation. Box offices charts were paradigmatic, but this period also saw a surge in interest in survey and focus group data, marketing white papers, metrics for tracking star image popularity, among others (Hayes and Bing 2004, 277–297). These media forms started on the margins, developed by relative outsiders, and operated quietly but powerfully in the background (as all functioning infrastructure does). They were either proprietary and hidden from public view or, as with box office charts, appeared to be straightforwardly reporting facts. They thus embody a strange contradiction in 1970s Hollywood: a flowering of artistry and auteur cinema accompanied by rather more mercenary ideas about how to reduce cinema to data points that could be tracked and analyzed by a consolidated suite of metrics.

What's in a Chart?

Like all lists, box office charts are deceptively simple. Their clean borders and ordered appearance seem simply to collate and present "objective" data. The form seems to repel critiques of its logic and instead demands judgements about value or importance based solely on weekly grosses. Unlike a piece of narrative prose or a scholarly essay, in which authorial voice, intent, or argumentation are discernable, a box office chart lacks such obvious traces. But as theorists of writing have shown, all acts of inscription, whether "administrative" or "literary," are the product of choices – what to include, or not, and how – that have wide-ranging consequences.[5] Though they appear simply to *gather* data, *read* trends, or *reflect* cultural activity that exists *a priori* in the world, such forms actually generate data or knowledge, give it material form, and grant it epistemological authority. Put another way, to measure a cultural field in a list or a chart makes concrete what had previously been abstract. It is to visualize data points, combine them into relations, attach metadata, and aggregate these into a standardized format.[6] For readers, the threshold to understanding such forms is low. They are easy to access, easy to scan, and thus easy to circulate as knowledge. Popular music charts, for instance, as I argue elsewhere (Young 2017), render songs, artists, moments, and memories into collective archives and canons, track commodity circulation, structure processes of taste-making and gatekeeping, and, in self-authored charts, allow for the performance of knowledge or mastery of a field. Box office charts function similarly. They draw things together and place them in relation to one another, creating connections that did not exist prior to the act of listing. They congeal as knowledge various components: films, box office earnings, studios, number of weeks on the charts. They create a stable format that is the material basis for conceptions of value, success, or failure. Consider the 1972 chart from *Variety* magazine shown in figure 1.

The basic criterion for this list is the total number of box office dollars earned per film over a given period of time (one week). Financial gross is privileged. All other data become metadata of secondary importance, such as distributor ("Distr."), last week's gross and rank, number of cities in which each film is screened, type of theatre (first run, show case, roadshow), number of weeks on chart, and total grosses to date. When presented in this way, such forms achieve a veneer of objectivity. Mary Poovey (1998) traces this sense of authority and objectivity back to early modern practices of double-entry bookkeeping. This technique of accounting, she argues, made concrete, on the page, previously abstract notions of transparency, trust, and even virtue before God. Things, events, amounts, and accounts are transcoded from speech and embodied experience into entries and numbers for elegant and efficient presentation. It's all there on the page, with

5 Many of these insights were derived during what are now referred to as the "orality-literacy" debates of the mid-20th century. Ong (1983) synthesizes much of this research, but earlier touchstone texts include Parry (1930), Parry and Parry (1987), Lord (2000 [1960]), Levi-Strauss (1966 [1962]), McLuhan (1962), Goody and Watt (1963), Havelock (1963), and Goody (1986).

6 For an expansive consideration of "format" as a critical concept, see Jancovic, Volmar, and Schneider, eds. (2020)

nowhere to hide and thus no reason to doubt the trustworthiness of the merchant with well-kept books.

Figure 1: Variety Chart 1972

Wednesday, October 25, 1972 VARIETY PICTURE GROSSES 17

50 Top-Grossing Films

[*WEEK ENDING OCTOBER 18*]

Compiled by Standard Data Corp., N.Y.

TITLE	DISTR	THIS WEEK $	THIS WEEK RANK	LAST WEEK $	LAST WEEK RANK	TOTALS CITIES	TOTALS FIRST RUN	TOTALS SHOW CASE	TOTALS ROAD SHOW	TOTALS THE-ATRES	WEEKS ON CHART	TOTAL TO DATE $
THE NEW CENTURIONS	COL	429,950	1	622,550	1	10	10	58		68	11	3,554,906
DELIVERANCE	WB	411,814	2	460,208	2	15	16	5		21	11	1,465,237
SUPER FLY	WB	299,200	3	421,500	3	16	18	4		22	11	4,509,174
FUNNY GIRL	COL	274,500	4	10,900		3	1	45		46	84	19,883,024
2001 SPACE ODYSSEY	MGM	187,300	5	37,625	24	5	4	30		34	135	15,944,342
COUNTESS DRCULA/VAMPIRE CIRCUS	FOX	179,000	6	32,000	29	2	1	36		37	2	211,000
BORN BLACK	HAL	175,000	7			1		33		33	1	183,340
CANCEL MY RESERVATION	WB	155,150	8	191,400	6	5	3	10		13	4	642,482
CABARET	AA	147,900	9	236,000	4	5	3	27	1	31	34	10,000,873
THE CANDIDATE	WB	144,200	10	213,250	5	4	2	28		30	16	2,394,492
EVERYTHING TO KNOW ABOUT SEX	UA	117,043	11	136,441	9	13	15	2		17	10	2,992,650
FIDDLER ON THE ROOF	UA	95,650	12	99,300	13	12			12	12	50	15,356,817
WHERE DOES IT HURT	CRC	90,450	13	165,350	7	8	11	4		15	10	1,001,793
BUTTERFLIES ARE FREE	COL	80,800	14	63,400	16	9	7	8		15	15	4,787,285
SOUNDER	FOX	80,629	15	34,676	28	3	4			4	3	159,065
HEAT	L-P	80,500	16	28,000	31	3	4	2		6	2	104,149
YOUNG WINSTON	COL	77,000	17			3			3	3	1	99,035
LAST HOUSE ON LEFT	HAL	73,000	18			3	3	7		10	4	151,200
FAT CITY	COL	71,100	19	37,900	23	7	8	11		19	11	717,760
THE RULING CLASS	AVE	69,700	20	90,750	14	6	6	2		8	5	247,293
ASYLUM	CRC	68,834	21			2	1	5		6	1	68,834
DR ZHIVAGO	MGM	62,800	22	162,275	8	5	5	9		14	32	2,840,730
THE GODFATHER	PAR	60,950	23	83,475	15	7	6	16		22	31	42,988,242
SLAUGHTERHOUSE FIVE	U	54,900	24	50,300	19	10	6	8		14	30	2,031,864
AIRPORT	U	50,000	25			1		26		26	50	13,214,642
ANDROMEDA STRAIN	U	50,000	26			1		26		26	24	3,552,415
THE MAN	PAR	48,000	27			1		19		19	11	775,998
MELINDA	MGM	45,540	28	47,000	20	4	4	2		6	9	1,837,478
MARJOE	C5	44,000	29	55,100	18	10	10			10	11	602,261
SECRET LIVES OF ROMEO & JULIET	AQU	41,000	30			1		26		26	1	88,900
NOTORIOUS CLEOPATRA	BI	41,000	31			1		26		26	1	84,700
BLUEBEARD	CRC	39,500	32	4,750		3	4	3		7	8	1,275,111
THE VIRGIN WITCH	BRE	36,000	33	5,500		1		10		10	5	210,700
HAMMER	UA	34,000	34	38,500	22	5	5			5	5	187,708
HICKEY & BOGGS	UA	31,750	35	127,000	10	3	2	10		12	7	588,397
EASY VIRTUE	VOG	31,400	36			2	1	4		5	7	289,470
THE EMIGRANTS	WB	31,278	37	35,421	25	2	2			2	3	87,104
FAREWELL UNCLE TOM	RGU	25,000	38			1	1			1	1	25,000
DEEP THROAT	AQU	23,675	39	25,000	34	1	1			1	18	529,750
SLAUGHTER	AIP	22,000	40	43,600	21	2	2			2	9	1,816,113
CHLOE IN THE AFTERNOON	COL	22,000	41	25,000	35	1	1			1	3	70,330
SORROW AND THE PITY	C5	21,500	42	18,500	39	5	5			5	11	428,979
ALL ABOUT SEX OF ALL NATIONS	MIS	21,283	43	25,260	32	1	1			1	35	1,053,861
GONE WITH THE WIND	MGM	20,000	44	11,375		3	3			3	72	10,672,071
WHEN LEGENDS DIE	FOX	20,000	45	16,400	42	1	1	3		4	4	88,800
A SEPARATE PEACE	PAR	19,000	46	22,000	37	1	1			1	3	61,775
WHATS UP DOC	WB	18,300	47	20,300	38	4	5			5	32	8,606,777
THE THING WITH TWO HEADS	AIP	16,500	48	23,234	36	1	1			1	5	130,358
BARON BLOOD	AIP	16,500	49	18,500	40	1	1			1	2	35,000
THUNDERBALL / ONLY LIVE TWICE	UA	16,000	50	34,700	27	1	1			1	18	1,938,644
ALL OTHERS		610,892		884,606			151	48		199		1,425,327,550
GRAND TOTAL		$4,883,488		$4,659,046			337	553	16	906		$1,605,915,479

Source: *Variety*, 25 October 1972, p. 17

Box office and other cultural charts participate in this long history, benefitting from long-held assumptions about the transparency and objectivity of numeracy and list. The material properties of the above chart defer questions about how this money is counted, who counts it, and whether these figures are reliable. The chart is instead suffused with a logic that privileges economic considerations. Its collation and hierarchical presentation of data seem obvious and inevitable. But this chart could have been organized much differently. One could have listed the number of tickets sold, rather than dollar amounts. One could have ranked entries according to number of screenings, or theatres in which they appeared. Similarly, one could have listed these films alphabetically, rather than according to total gross amount, or chronologically based on date of release. There are innumerable ways of ordering and reordering such data (this is the power and pleasure of any list), and innumerable metadata that could be affixed to broaden or reduce the scope of the list. We do not see, for instance, the country of origin, name of director or lead actor, number of craftspeople who worked on the film, years since initial script development, and so on. These may today be considered irrelevant according to industry standards, but such judgements ignore the extent to which these industry standards are themselves established through informational forms such as box office charts. Because alternative data were excluded from early charts, or presented as contextual metadata, now they always will be.

These techniques of assembling data recall Bruno Latour's early science studies research, in which he describes the "captation" of readers who are beholden to a particular pathway charted by the assembler(s) of a document through an archive of almost infinite possible connections and pathways. Readers are held captive, their objections anticipated and neutralized, while the objects and relations assembled come to be dominated by sight. Words, things, people, and events are frozen in visual forms such as lists, tables, charts, or diagrams in order that they can better be controlled from a distance, as he writes, "when someone is said to 'master' a question or to 'dominate' a subject, you should normally look for the flat surface that enables mastery (a map, a list, a file, a census, the wall of a gallery, a card-index, a repertory) and you will find it" (Latour 1990, 45). Knowledge is formed via these slow and sedimentary processes by which many networks of observation, experimentation, commentary, citation, and classification are coordinated and stabilized. People do things with such knowledge. They use it to analyse situations and make decisions, distribute resources, and to otherwise coordinate action. Theories of historical rupture make for good headlines and citation counts, but the process of history is slow, layered, and messy.

Box office charts, for instance, establish particular rhythms of time over film production, circulation, and reception. As Hayes and Bing demonstrate, for at least the last thirty years, a film's opening weekend gross has stood as the gold standard of success (Hayes and Bing 2004, 9–15). Prior to the standardization of box office charts in the 1970s, Hollywood release and exhibition practices did not always pass through this temporal bottleneck. Early *Variety* box office reporting "didn't publish a single regular report on nationwide grosses. Instead, the paper ran a bewildering array of regional stories reporting ticket sales at particular houses" (Hayes and Bing 2004, 290). Eventually, in-house critic A.D. Murphy began to aggregate these data to derive rudimentary national numbers based on averages from key regions in a column he began to author in 1968, "*Variety*'s Key City B.O. Sample." As the 1970s unfolded, Murphy began to zero in on first week grosses and by the early 1970s, his charts had taken shape in the now-familiar "top weekend grosses" format (Murphy 1968, 290–291). More than mere "information," box office charts became operational and logistical media, powerful switch points in the feedback loops between Hollywood film production, distribution, exhibition, and consumption. Hollywood executives lusted after total data awareness of film performance and audience tendencies, joining military commanders and 20th century Cold Warriors in search of seamless connectivity and real time data, or retail executives like Sam Walton, founder of Wal-Mart, who sought ever more granular data about logistics and distribution (LeCavalier 2016). The payload at the end of each data supply chain varies, but the goals are the same: speed, efficiency, and optimization. New intermediaries arrived in Hollywood to fill this niche, convinced that they could develop "a more scientific system for building a movie into a mass event, and a more scientific system for rapidly measuring its success" (Hayes and Bing 2004, 241).

Data Imaginaries

Murphy was not the only quantitative game in town. As Hayes and Bing (2004, 283–296) recount, other groups like National Research Group (NRG) and Centralized Grosses were attempting to streamline and aggregate data about the business of Hollywood. NRG was founded in 1979 by two former political pollsters, Catherine Paura and Joe Farrell, to provide Hollywood studio executives with market research and other curated data sets that, they argued, offered unique insights into what was popular, how, and why. A few years earlier, in 1976, Marcie Polier, then a part-time secretary for Mann Theatres in Los Angeles, founded Centralized Grosses. At Mann, Polier canvassed individual theatre owners about ticket sales and collated their data into a report for distribution to industry executives. Centralized Grosses systematized what had been an *ad hoc* and relatively unreliable process of data collection (based on a complex of voice – telephone – pencil – typewriter – binder). It also transformed Polier's early local data sets into a nationwide database that enjoyed privileged access to proprietary studio data for over two decades. The successes of NRG and Centralized Grosses exemplify the datafication of culture that Murphy's charts similarly embody. They also stand as early examples of broader transformations in Hollywood arising from ownership concentration and the logic of financialization, which Andrew deWaard describes as an "accelerated growth of the financial sector and its extractive logic, which relies on financial engineering rather than commodity production" (deWaard 2020, 54–55). Polier sold her company, for instance, to AC Nielson, the notorious global market research and television rating system corporation, before it was merged, acquired, and stripped for parts by a series of media companies and private equity firms. The initial sale of Centralized Grosses (by-then rebranded as Entertainment Data, Inc.) to Nielson resulted in a new company, Nielson EDI. The latter was purchased in 2002 by Dutch publishing company VNU, which then owned iconic trade papers *Adweek, Billboard,* and *Hollywood Reporter*. In 2006, a consortium of private equity firms purchased VNU, installed the former CEO of General Electric, and rebranded the entity as The Nielson Company. As deWaard shows, the consortium saddled the company with excessive debt, stripped its assets for capital extraction and slashed its workforce before exiting the investment with a handsome profit "achieved through financial engineering" (deWaard 2020, 67). The journey of Polier's Centralized Grosses shows how data practices become capitalized and financialized, but it also reminds us that the roots of the data economy lay in clerical labor and bookkeeping, areas of work often feminized, racialized, and ignored in business and cultural histories.

Murphy's chart operations at *Variety* are of particular note for our purposes given their origins in the military-industrial complex (the connection to military-style systems thinking I suggested above is more than a provocative analogy). Before landing at *Daily Variety* in 1965, he had trained as a navigator and transporter in the United States Navy and written a Master of Science thesis at the United States Naval Postgraduate School on "A Queueing Model of Information Flow in a Command and Control System" (Murphy 1962). These intellectual leanings in fact played a role in his ability to secure a job as a conventional film critic at the magazine. He wrote a letter in 1965 to editor Tom Pryor that modeled a mathematical, seemingly objective analysis of the economics of Hollywood that caused Pryor to hire him immediately (Hayes and Bing 2004, 289–290). Murphy's peculiar background is often mentioned in histories of *Variety* and the box office charts, yet rarely analyzed in depth. His Master's thesis, publicly available, offers clear insight into the types of knowledge that would shape his approach to working with data and systematizing the charts. Though much of the analysis is highly technical, Murphy's opening and concluding chapters lucidly lay out the argument and its epistemological commitments. He explores the speed, volume, and processes by which information moves in military chains of command, seeking to address the problem of military commanders having to process too much information. In his own words, the thesis seeks to "provide a methodology yielding quantitative results which may assist a commander" in addressing "the problem of the volume of information flow" (Murphy 1962, ii). "Does a military commander actually need all the information he receives? Can he afford to spend more and more time absorbing all the subtleties of increasingly complex matters, matters on which he must make positive decisions and decisions for which he alone is responsible? Can he afford NOT to?" (Murphy 1962, iii). The basic premise of the thesis is that decision making can be optimized by streamlining and standardizing the movement of information from one command post or "service stand" to the next (Murphy 1962, v). It is unclear why Murphy adopts the term *information*, as opposed to *knowledge* or *data*, but it likely had something to do with the spread through the 1950s and 60s of Claude Shannon's so-called *information theory*. Though he does not cite or comment upon Shannon directly, nor on Norbert Wiener's cybernetics, the influence of both is clearly evident in Murphy's use of the *information* and *command and control* concepts throughout.

These nascent fields offered modes of analysis and reasoning employed by Murphy in suggesting that there are informational thresholds beyond which any given command post ceases to be functional, and that these thresholds can be quantified. Short of such a threshold, he argued, any actor's capacity for reasoning remains optimal and she or he can devote maximum cognitive capacity toward complex reasoning, analysis, and decision making that is required in a given situation. Murphy pointed to protocols such as Standard Operating Procedures (SOP) as evidence of how such practices were already in place. Whether in military or institutional settings, SOPs are systems designed with the goals of reducing uncertainty, increasing redundancy, and lowering the threshold to participation in any chain of decision, command, or data transfer. SOPs program action according to anticipated parameters and scenarios and thus reduce the cognitive load on decision makers and accelerate their ability to act. One need not reason nor analyze one's way toward a decision or action, one must simply adopt the SOP. The critique of instrumentalizing and rationalizing decisions involving, for instance, weaponry and human life is obvious. But one can also understand the desire, in situations that demand speed and "decisiveness," to outsource some of the cognitive load.

Murphy was a fan of SOPs but pointed out that even most military SOPs remained *ad hoc* and without an important quantitative dimension that could make them even more efficient, functional, and standardizable. His hunch was that a synthesis of cybernetics' command and control systems analysis with the hard mathematics of queuing theory could address these limitations. Ultimately, Murphy's solution to the problem of information overload in decision-making is to call for comprehensive analyses of chains of command and information transfer "with a view to automating, consolidating, or even eliminating some of the information generated" (Murphy 1962, 34) so as to lessen the cognitive load of commanders. In other words, to streamline, optimize, and eliminate redundancy should be, according to Murphy, the primary goals for any such system.

Though Murphy aspires to general applicability with his analysis, he lacks any theory of what the analysis models, namely, *communication* and *information*. There is no conception of the devices, agents, or practices that perform the work of communication, that mediate the flows of information he is so concerned with. Murphy instead relies on a concept of information, common since the mid-19$^{\text{th}}$ century but popularized through misreadings of Claude Shannon and Norbert Wiener, that mistakes their entropy measure, information, for a universal and ahistorical substance "out there" in the world to be discovered, captured, nurtured, saved, protected, or otherwise utilized. Such a conception ignores how practices, techniques, models, mechanisms, and media do not simply reflect or gather but actively generate data points, processing them into *information* to be disseminated as *knowledge* for various purposes. As Gitelman et al. (2013) show, following Bowker (2005), "raw data is an oxymoron." In attempting to derive general principles that can be used in any number of different communicative environments, Murphy's model assumes (or dreams) of a frictionless set of processes by which data are discovered or pulled from "the field," passed to a decision maker, processed cognitively into a set of decisions, conveyed as orders to actors downstream, and implemented in the field.

Murphy's thesis is yet another instance in which *information* takes on almost magical properties, as Nunberg writes, "a noble substance [...] indifferent to the transformation of its vehicles" and to the techniques of assembly, movement, and operationalization associated with information work (Nunberg 1996, 107). This way of observing a communicative and logistical field describes, more or less, Murphy's eventual approach to charting and analyzing the contours of Hollywood's film industry. Once established at *Variety,* he began to ask: how could this noisy information environment – riven with rumors, innuendo, gut feelings, and the uncritical acceptance of conventional wisdom – be made more efficient for decision makers? How could flows of information, capital, and labor be optimized for expansion and growth? Could quantitative methods and mathematics be used to clear Hollywood of its emphasis on aesthetic and unscientific concerns? His *Variety* charts comprise a longitudinal and consequential answer to such questions. Box office charts should be thus read as a remediation of data practices that were central to military imaginaries of the 20th century.

Hollywood Time

One of the primary avenues by which box office charts standardized and optimized the business of Hollywood was in establishing particular temporal rhythms. As mentioned above, before it was systematized into a formal chart system, box office reporting was organized primarily around spatial considerations – region, city, district, or theatre. But as "this week," "previous week," and "weeks on chart" became increasingly ubiquitous in trade and newspapers, a new set of temporal desires began to emerge among industry executives and financiers. Charts satisfied these desires, which they had helped create. Long-standing head of Universal, Lev Wasserman, for instance, became throughout the 1970s obsessed with obtaining real time updates on as many aspects of Universal's operations as possible, demanding, according to biographer Connie Bruck, "slips of paper [delivered] throughout the day with updated box office figures, the Universal [Studios] Tour head count, and stock market closing numbers" (quoted in Hayes and Bing 2004, 293). The charts helped quench the thirst for the latest figures. Their formal emphasis on the week and weekend eventually consolidated almost all industry attention and resources toward these as the primary units of a film's lifecycle. But their formal attributes had subtler implications on "Hollywood time." Charts function as shared archives that freeze for posterity a snapshot of what is popular in a given moment, inscribing a perpetual forward momentum on a cultural field. The "previous position" column frames the present week as part of a longer trajectory, either rising or falling, and the reader – whether industry insider or casual film fan – always has the next week and future chart performance in mind.

Such functions are an example of what Volmar and Stine (2021) describe as "hardwired temporalities." This concept accounts for enduring temporal patterns that are produced and sustained through infrastructures consisting of material artifacts or technologies, on the one hand, and different labor and other practices, on the other. Box office charts are "paper machines" (Krajewski 2011) that hardwire particular temporal rhythms through the slow and often painstaking work of assembling, organizing, and displaying data. In such ways, box office charts function similarly to pop music charts, which, according to Straw (2015), "transform the often erratic commercial life of a musical commodity into a curve of ascendant and descendant popularity, so as to endow that life with the legibility of both narrative and tabular form" (129). This curve informs the reader's judgement of an entry's performance (successful or not), relevance (popular or not), or value (good or bad), eliciting speculation about its future – is it headed toward #1, or will it be pushed off the chart's bottom edge? Many times converge and become standardized according to such a chart's logic of competition: collated pasts, anticipated futures, and present popularity. Charts freeze the present as part of an ongoing archiving of cultural history, creating what Hakanen calls a "stockpile" of information that can be used for analysis, prediction, and narrativization (Hakanen 1998, 106). According to Straw, the curve of an entry's "rise and fall endows its lifecycle with the romance of individual success and ultimate exhaustion, while the hierarchical verticality of the chart conveys the sense of all songs sitting in momentarily stable relationships within a homogenous historical moment" (Straw 2015, 132). Time is rendered spatial, materialized as data points and displayed on paper. Insiders can base marketing decisions that affect future film performance on such data, while fans can rely on the charts to provide knowledge about films they have not yet seen. The bias of charts is toward organizing past events hierarchically and predicting future results for the purposes of modelling and decision making. These logistical functions are in constant tension with the appearance of charts as a finite and disposable form. The momentary order of a chart will dissolve and be replaced by the next chart, which is always on the horizon. "The charts" go on, but this chart will not. Charts fetishize newness because they never allow consumers to stop looking at the present in terms of the future.

Charts render time spatial, flattening the rhythms of commodity circulation and consumer desire into pre-formatted categories in two dimensions (last week, this week, next week). This process is mirrored in the ways charts flatten difference and aesthetic form. In a box office chart, the film-as-artwork becomes datafied, rendered into units that can be sorted and recombined in various ways. When abstracted and fitted to the templates of a chart, films become standardized and thus more suitable for direct competition. More traditional criteria used to differentiate films, such as genre, geographical provenance, style, or audience disappear on the charts. This flattening of difference ensures films and artists must compete with one another for chart, and thus market, supremacy. As with pop music charts, "[w]ithin the market-place there is only competition on the basis of assumed equivalence; any differences are reduced to differential calculations about exchange value" (Parker 1991, 211). The logic and values of capital are encoded in the structural attributes of box office charts. *Art Murphy's Box Office Register* exemplifies this market-based logic. In the introduction to 1985's edition, he writes,

> Box office grosses must be looked at constantly – on almost EVERY film, on almost EVERY day. Far from being mere numbers, box office grosses represent the responses of PEOPLE to films. To ignore those numbers – and those people – is to risk business failure and – worse – to inhabit the catatonic world of the compulsive aesthete (quoted in Hayes and Bing 2004, 291).

Murphy's charts were clearly suffused with a logic of efficiency and optimization, begging the question of whom or what this optimization serves. In this case, the question of politics is straightforward. Murphy was virulently anti-studio and antilabor. He viewed Hollywood's unions as part of the system rather than an oppositional force. He railed against the combined hegemony of the studios and unions, such as in a 1968 op-ed in *The New York Times* that warned young American creative talent about "inbred, protective unionism [that] has erected a formidable barrier to your entry into Hollywood movie making." Studio management, he continued, was only too happy to go along with the current state of affairs given its "congenital" fear of labor unrest. Not only this, he lamented, but "[a]stonishingly, an agency of the United States Government actually is helping to perpetuate closed union ranks, and thereby suppress any professional competition for work" (Murphy 1968, 7). Murphy continued to wear these conservative inclinations on his sleeve until the end of his life. As John Welles, a highly successful Hollywood producer and former student of Murphy's, wrote upon the latter's death, "I'll miss him calling me every Thursday morning to [gripe] about the liberal politics of *The West Wing*" (quoted in Lapriore 2003). Though critical of studio oligopolies, it's clear that Murphy's notion of optimizing the business of Hollywood was in the service of capital, not labor. It is in this context that we should situate his *Box Office Register*. Though he claimed it was not more than a compendium of "numbers which represent the collective and comparative decisions of paying customers," a "convenient [and] cohesive understanding what makes moving pictures move" (quoted in Hayes and Bing 2004, 291), the *Registry* was much more. It was a moral argument about efficiency and value extraction, an attempt to rescue the business of Hollywood from the whims of the artists and what he called "compulsive aesthetes" (quoted in Hayes and Bing 2004, 291). Such rhetoric would have been familiar to readers amid the rising neoliberal tide of Reagan and Thatcher in 1980s, and well into the deregulation and globalization of the 1990s.

Murphy's influence on Hollywood as a cultural industry went still further. As the charts continued to spread throughout the culture, he turned his eye to teaching, developing a graduate seminar on the business of film in 1974 for the University of Southern California's film school, which he taught until 1997. The success of this course allowed Murphy to play an active role in designing USC's Peter Stark Program, for which he served as director from its inception in 1979 until 1990, and in which he taught until 1997. This began as a general program in film business but evolved to become focused on the role of film Producer. Through the 1980s and 90s, this figure, an intermediary between studio management, financiers, and creative talent, would help erode the power and influence of the auteur-director and even the star performer, many of whom don't truly "arrive" as Hollywood megastars until they establish their own production companies. In his role at USC, Murphy mentored a generation of students that would move into film production with an eye for his modes of measuring performance and assessing value, typified by his charts and the *Register*, that could be fed back into the production process to inform decisions about marketing and resource distribution. Given the centrality of Murphy to the Stark program's development and delivery, particularly in its early years, it seems likely that many of these epistemological assumptions would have found their way into the program's curriculum. This remains an open question as any such records have either been lost or remain inaccessible to outside researchers.[7] In any case, the Stark program was emblematic of a shift within Hollywood, away from aesthetics and toward corporate logics, that corresponded with the changing epistemological ground as box office charts and other informational forms took hold over the industry's collective imagination.

This shift, the datafication of culture in the name of market analysis, merges ideas of artistic or creative "quality" with the rate and speed of financial performance. Assessment of Hollywood films now almost always passes through the bot-

7 I was told in personal correspondence by the current director of the program that the program has not kept records of past syllabi or other documentation related to curriculum development, at least none that are accessible to outside researchers.

tleneck of chart performance. The days of films slowly building up momentum through word of mouth, favorable reviews from critics, or redemptive readings from "cult" audiences, such as were commonplace even into the 1990s, have been steadily eroded by financial and other performance-based metrics. As Hayes and Bing (2004) continuously point out, the marketing emphasis on any given film is aimed entirely at its opening weekend and thereafter at keeping it at or near the top of the box office chart. If a film "flops" its opening weekend, support and resources are quickly reallocated elsewhere in the studio's repertoire and the film disappears from view almost immediately. The compression of these time horizons corresponds with the synthesis of box office chart and film review on aggregator websites such as RottenTomatoes.com or Metacritic.com. These websites not only list box office performance on their home pages but also port the quantitative logic of the charts into the realm of criticism, creating an aggregated "review score" that reduces an array of film reviews and ratings to a single number. Though Hollywood insiders lament this development, such as the parade of anonymous sources in Barnes' (2017) feature on Rotten Tomatoes in *The New York Times*, it's clear that aggregators are top-of-mind for studio financiers and marketers – a fact only underscored when considering that RottenTomatoes is now owned by two studios (Fandango, a subsidiary of NBCUniversal, owns 75%; WarnerMedia owns 25%). It falls beyond the scope of this essay to explore such contemporary developments, but connections are suggestive.

Conclusion

This chapter has sought to contextualize Hollywood's contemporary privileging of Big Data and financialization within a longer historical arc of data practices and techniques of counting culture. Hollywood box office charts are a paradigmatic example of how practices of measuring and counting that informationalize and "datafy" commodities and behavior have cascading implications in the cultural industries. Films as aesthetic objects and, by extension, filmmaking as a complex of collaborative practices, become abstracted into categories of metadata deemed relevant by chart and industry insiders, visualized and flattened for inscription on a two-dimensional surface or interface, then ranked hierarchically by a single metric (box office gross over a given unit of time). These charts are easy to reproduce and thus propagate widely across media environments. They are visually appealing against the walls of text that typically surround them. And they are accessible for a wide audience given that they do not demand close reading but rather scanning and inspection.

Box office charts are a species of *logistical media* in the way they coordinate action and establish or reinforce infrastructures of space and time for the film industry. New rhythms of time were inscribed on box office charts that pushed the attention, resources, and emphasis onto the opening week and weekend as the primary unit by which to track, measure, and assess any given film's success or failure. And as charts spread across the culture, insiders operationalized them in decision-making processes that affected film production at every moment along the supply chain, from development to exhibition and beyond. Such informational forms, and the data practices that produced them, became an increasingly important zone of competition – not just between films contained in the charts but among studio heads looking for any advantage, informational, economic, or otherwise, over their competitors.

While box office charts are precursors of contemporary recommendation, search, and newsfeed algorithms, the latter have clearly exceeded the scope and scale of box office charts and similar "paper machines." In digital environments, the granularity of measurement, speed of aggregation, complexity of calculation, and specificity of outputs individually tailored to each platform user, all vastly exceed what early chart producers such as A.D. Murphy could have imagined. These cloud-based data practices of audience measurement and tracking have wide-ranging implications on the production, circulation, reception of culture, as examined elsewhere in this volume. My hope is that the above analysis shows that differences between digital and paper-based data practices are of degree, not kind. Looking to these earlier cultural moments and forms allows media and cultural historians to observe inflection points, moments in which the epistemological ground begins to shift. Because these early data practices were not yet widespread, and often occurred in single publications, they offer the benefit of analytic and methodological clarity. They provide resources for reconstructing genealogies of cultural activity, while also deflating narratives emanating from Silicon Valley that fetishize the new-ness of digital culture.

Like most logistical media, box office charts have become so familiar that it is difficult to bring their operations into view. The more familiar an infrastructure or a data practice becomes, the easier it is to assume its universality and ahistoricism, as if it was inevitable or eternal. Scholars have lately become more attuned to these questions, highlighting that all structures of knowledge and all data practices are historically contingent, subject to power asymmetries, and open to revision. As we spend more of our time enframed and thus perplexed by "consumption junctions" (Greenberg 2008; Vonderau 2015) like aggregators, algorithms, recommendations, and play-lists, we might cast our eyes back to learn from the data practices that mediated similar struggles over resources, representation, and epistemology in previous historical moments and media environments.

References

Acland, Charles R. 2020. *American Blockbuster: Movies, Technology, and Wonder*. Durham, NC: Duke University Press.

Barnes, Brooks. 2017. "Attacked by Rotten Tomatoes." *The New York Times*, September 7. Accessed August 26, 2021. https://www.nytimes.com/2017/09/07/business/media/rotten-tomatoes-box-office.html.

Biskind, Peter. 1999. *Easy Riders Raging Bulls: How the Sex-Drugs-And Rock'N Roll Generation Saved Hollywood*. New York City: Simon and Schuster.

Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

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

Buckland, Warren, and Christopher Long. 2006. "Following the Money: The Lord of the Rings and the Culture of Box Office Figures." In *The Lord of the Rings: Popular Culture in Global Context*, edited by Ernest Mathjis, 88–98. London: Wallflower Press.

Case, Judd A. 2013. "Logistical Media: Fragments from Radar's Prehistory." *Canadian Journal of Communication* 38 (3): 379–395.

deWaard, Andrew. 2020. "Financialized Hollywood: Institutional Investment, Venture Capital, and Private Equity in the Film and Television Industry." *JCMS: Journal of Cinema and Media Studies* 59 (4): 54–84. Accessed August 26, 2021.

Eco, Umberto. 2009. *The Infinity of Lists*. New York: Rizzoli.

Elsaesser, Thomas, Alexander Horwath, and Noel King, eds. 2004. *The Last Great American Picture Show*. Amsterdam: Amsterdam University Press.

Finn, Jonathan. 2016. "Timing and Imaging Evidence in Sport: Objectivity, Intervention, and the Limits of Technology." *Journal of Sport and Social Issues* 40 (6): 459–476.

Frohmann, Bernd. 2004a. "Documentation Redux: Prolegomenon to (Another) Philosophy of Information." *Library Trends* 52 (3): 387–407.

Frohmann, Bernd. 2004b. *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.

Gitelman, Lisa, ed. 2013. *"Raw Data" Is an Oxymoron*. Cambridge, MA: MIT Press.

Goody, Jack. 1977. *The Domestication of the Savage Mind*. Cambridge, UK: Cambridge University Press.

Goody, Jack. 1986. *The Logic of Writing and the Organization of Society*. Cambridge, MA: Cambridge University Press.

Goody, Jack, and Ian Watt. 1963. "The Consequences of Literacy." *Comparative Studies in Society and History* 5 (3): 304–45.

Greenberg, Joshua M. 2008. *From Betamax to Blockbuster: Video Stores and the Invention of Movies on Video*. Cambridge, MA: MIT Press.

Guillory, John. 2004. "The Memo and Modernity." *Critical Inquiry* 31 (1): 108–132.

Hakanen, Ernest A. 1998. "Counting down to Number One: The Evolution of the Meaning of Popular Music Charts." *Popular Music* 17 (1): 95–111.

Havelock, Eric A. 1963. *Preface to Plato*. Cambridge, MA: Harvard University Press.

Hayes, Dade, and Jonathan Bing. 2004. *Open Wide: How Hollywood Box Office Became a National Obsession*. New York: Miramax Books.

Hayward, Mark. 2019. *Identity and Industry: Making Media Multicultural in Canada*. Montreal: McGill-Queen's University Press.

Huber, Alison. 2010. "Making Time Stand Still: How to 'Fix' the Transient Top 40." *International Journal of Cultural Studies* 13 (2): 147–162.

Jancovic, Marek, Axel Volmar, and Alexandra Schneider, eds. 2020. *Format Matters: Standards, Practices, and Politics in Media Cultures*. Lüneburg: Meson Press.

Kittler, Friedrich. 2010. *Optical Media*. Cambridge: Polity.

Kittler, Friedrich A. 1999. *Gramophone, Film, Typewriter*. Translated by Geoffrey Winthrop-Young and Michael Wutz. Palo Alto, CA: Stanford University Press.

Krajewski, Markus. 2011. *Paper Machines: About Cards & Catalogs, 1548–1929*. Cambridge, MA: MIT Press.

Lapriore, Elaine. 2003. "Passings." *USC News*, June 18. Accessed August 27, 2021. h ttps://news.usc.edu/1581/Passings/.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1990. "Drawing Things Together." In *Representation in Scientific Practice*, edited by Michael Lynch and Steve Woolgar, 19–68. Cambridge, MA: MIT Press.

Lauer, Josh. 2017. *Creditworthy: A History of Consumer Surveillance and Financial Identity in America*. New York City: Columbia University Press.

LeCavalier, Jesse. 2016. *The Rule of Logistics: Walmart and the Architecture of Fulfillment*. Minneapolis: University of Minnesota Press.

Levi-Strauss, Claude. 1966 [1962]. *The Savage Mind*. Chicago: University of Chicago Press.

Lord, Albert Bates. 2000 [1960]. *The Singer of Tales*. Edited by Stephen Arthur Mitchell and Gregory Nagy. Cambridge, MA: Harvard University Press.

Mathijs, Ernest. 2006. *The Lord of the Rings: Popular Culture in Global Context*. New York City: Wallflower Press.

Mattern, Shannon. 2013. "Methodolatry and the Art of Measure." *Places Journal*, November. Accessed August 27, 2021.

McLuhan, Marshall. 1962. *The Gutenberg Galaxy*. Toronto: University of Toronto Press.

Murphy, A.D. 1962. "A Queueing Model of Information Flow in a Command and Control System." M.Sc. Thesis, Monterey, CA: United States Naval Postgraduate School.

Murphy, A. D. 1968. “Students: Stay out of Hollywood.” *Cinéaste* 2 (6): 7–27.

Nunberg, Geoffrey, ed. 1996. *The Future of the Book*. Oakland: University of California Press.

Nyckel, Eva-Maria, and Nikolaus Poechhacker. 2020. “Logistics of Probability: Anticipatory Shipping and the Production of Markets.” In *Explorations in Digital Cultures*, edited by Marcus Bukhardt and Mary Shnayien. Lüneburg: Meson Press.

Parker, Martin. 1991. “Reading the Charts – Making Sense with the Hit Parade.” *Popular Music* 10 (2): 205–217.

Parry, Milman. 1930. “Studies in the Epic Technique of Oral Verse-Making. I. Homer and Homeric Style.” *Harvard Studies in Classical Philology* 41: 73–147.

Parry, Milman, and Adam Parry. 1987. *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford: Oxford University Press.

Peters, John Durham. 1988. “Information: Notes toward a Critical History.” *Journal of Communication Inquiry* 12 (2): 9–23.

Peters, John Durham. 2013. “Calendar, Clock, Tower.” In *Deus in Machina*, edited by Jeremy Stolow, 25–42. New York City: Fordham University Press.

Peters, John Durham. 2015. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press.

Phiddian, Robert, Julian Meyrick, Tully Barnett, and Richard Maltby. 2017. “Counting Culture to Death: An Australian Perspective on Culture Counts and Quality Metrics.” *Cultural Trends* 26 (2): 174–80.

Poovey, Mary. 1998. *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: University of Chicago Press.

Reeves, Joshua, and Jeremy Packer. 2013. “Police Media: The Governance of Territory, Speed, and Communication.” *Communication and Critical/Cultural Studies* 10 (4): 359–384.

Robertson, Craig. 2021. *The Filing Cabinet: A Vertical History of Information*. Minneapolis: University of Minnesota Press.

Rossiter, Ned. 2016. *Software, Infrastructure, Labor: A Media Theory of Logistical Nightmares*. London: Routledge.

Serres, Michel. 2007. *The Parasite* [1980]. Translated by Lawrence R. Sher. Minneapolis: University of Minnesota Press.

Shapiro, Aaron Murray. 2018. *Design, Control, Predict: Cultural Politics in the Actually Existing Smart City.* PhD Dissertation. University of Pennsylvania.

Siegert, Bernhard. 2015. *Cultural Techniques: Grids, Filters, Doors, and Other Articulations of the Real*. New York City: Fordham University Press.

Straw, Will. 2015. “Mediality and the Music Chart.” *SubStance* 44 (3): 128–138.

Swartz, Lana. 2020. *New Money-How Payment Became Social Media*. New Haven: Yale University Press.

Taschereau Mamers, Danielle. 2017. *Settler Colonial Ways of Seeing: Documentary Governance of Indigenous Life in Canada and Its Disruption.* PhD Dissertation. University of Western Ontario.

Thylstrup, Nanna, and Stina Teilmann-Lock. 2017. "The Transformative Power of the Thumbnail Image: Media Logistics and Infrastructural Aesthetics." *First Monday* 22 (10). Accessed August 27, 2021.

Virilio, Paul. 1989. *War and Cinema: The Logistics of Perception*. Brooklyn: Verso Books.

Virilio, Paul, and Benjamin H. Bratton. 2006. *Speed and Politics*. Los Angeles: Semiotext(e).

Vismann, Cornelia. 2008. *Files: Law and Media Technology*. Translated by Geoffrey Winthrop-Young. Stanford, CA: Stanford University Press.

Volmar, Axel. 2020. "Reformatting Media Studies: Toward a Theoretical Framework for Format Studies." In *Format Matters: Standards, Practices, and Politics in Media Cultures*, edited by Marek Jancovic, Axel Volmar, and Alexandra Schneider. Lüneburg: Meson Press.

Volmar, Axel and Kyle Stine, eds. 2021. *Media Infrastructures and the Politics of Digital Time. Essays on Hardwired Temporalities*. Amsterdam: Amsterdam University Press.

Vonderau, Patrick. 2015. "The Politics of Content Aggregation." *Television & New Media* 16 (8): 717–733.

Wajcman, Judy. 2019. "How Silicon Valley Sets Time." *New Media & Society* 21 (6): 1272–1289.

Werbin, Kenneth C. 2017. *The List Serves: Population Control and Power*. Amsterdam: Institute of Network Cultures.

Wernimont, Jacqueline. 2019. *Numbered Lives: Life and Death in Quantum Media*. Cambridge, MA: MIT Press.

Yates, JoAnne. 1993. *Control through Communication: The Rise of System in American Management*. Baltimore: John Hopkins University Press.

Young, Liam Cole. 2017. *List Cultures: Knowledge and Poetics from Mesopotamia to Buzzfeed*. Amsterdam: Amsterdam University Press.

II: Data Ethnography

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 Digital*. 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/10.1177/000276499 21955326.

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/10.1177/2053 951717720949.

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.

"Girls are like Glass": Situated Knowledges of Syrian Refugee Women on Datafication and Transparency

Araa Al Jaramani, Sandra Ponzanesi, and Gerwin van Schie

> Girls are like glass. They are easy to break or scratch, distorting their glazy nature. This metaphor about women in our society is something you hear everywhere in Syria: Girls are like glass! Behind and before this metaphor stand many carefully and purposefully untold stories, because telling them would mean "breaking the glass." [Mrs. B in Araa Al Jaramani]

The quote emerged during a three-day workshop, in which Mrs. B participated, on the importance of writing as a tool to tell human stories that Araa Al Jaramani, the first author of this chapter, organized for a women's group. While at first Mrs. B seemed to have no interest in the ongoing discussion, she suddenly came up with this remarkable statement. We used it as an epigraph because it encapsulates many of the discussions we tackle in this chapter on women's voice and position, issues of datafication and transparency, and the role of narrative in creating space for alternative subjectivities and identities in the condition of migration and integration.

In 2013, Araa Al Jaramani, the first author of this chapter, fled the war in Syria and accidentally ended up in the Netherlands after the human traffickers she paid to bring her to Sweden did not keep their promises. After arrival, Araa officially requested asylum and started her process with the Dutch Immigration and Naturalization Service (Immigratie- en Naturalisatiedienst, IND). Throughout the process, Araa was struck by its bureaucratic nature, the sometimes-amateurish conduct of the interviewers and interpreters, and the gender-insensitive approach to the very particular experiences of refugee women. After she was granted her refugee status, she started and led an organization to support Syrian women throughout their asylum-seeking and integration processes; she heard similar stories from these women being frustrated with their treatment by the IND. When given the opportunity during a postdoc project on Syrian refugee women as data subjects that she carried out at Utrecht University, Araa started to document her own experiences and the stories other women told her. In addition, together with Sandra Ponzanesi and Gerwin van Schie, co-authors of this chapter, she took another critical look at the data in the file that the IND collected about her, by studying the copy which was given to her during the asylum procedure. The copy of this file is given to each asylum seeker to provide a sense of transparency in the otherwise daunting and stressful time of awaiting the decision on whether or not they will be granted permission to stay. While this procedure does give any asylum seeker the opportunity to read this file and suggest corrections, in this chapter we will critically question which of the parties involved actually benefits from this "transparency," since, for a refugee fleeing from war, nothing is more opaque than the bureaucratic procedures of a foreign (from their perspective) country's agency whose mission it is to "limit and control refugee flows to Europe" (Ministry of General Affairs 2016, translation by authors).

Yet transparency is an often-heard mantra in a society that increasingly relies on data and algorithmic processes in its bureaucratic organizations (Lathrop and Ruma 2010; Hansen and Flyverbom 2015; Douglas and Meijer 2016). Organizations are increasingly becoming aware that they need to open up to consumers and citizens about their often-hidden digital processes, since many aspects of business and governance are becoming subject to some form of datafication (van Es and Schäfer 2017). Datafication is understood here as the principle that underpins the transformation of all available information, often about the conduct and behavior of people, into quantifiable units, making them available for access, monitoring, analysis, and prediction (Mayer-Schönberger and Cukier 2013; van Dijck 2014). In line with this larger societal trend, the IND too has implemented several systems that rely on processes of datafication in their bureaucratic practice. However, as we will show, operational transparency with regard to forced migrants in a situation of limbo, as neither not-yet-citizens nor never-citizens, seems to be rather one-directional as it mainly aids the bureaucratic and legal needs of the IND rather than the emotional and humanitarian needs of asylum seekers.

Recent research on data processes employed in forced migration policies has mainly focused on data practices by the European Union in order to protect "Fortress Europe" (Dijstelbloem and Meijer 2011; Leurs and Smets 2018), or use by Syrian refugees themselves of social media and the telephone (Sánchez-Querubín and Rogers 2018; Gillespie, Osseiran, and Cheesman 2018). Media and social research on immigration practices has largely concentrated on inequality and the representation of refugees in society (Schinkel 2013; Castles, Haas, and Miller 2013; Ticktin 2006), in policy-making (Rath 2001; van der Haar and Yanow 2011; Bakker, Cheung, and Phillimore 2016), and in the process of integration (De Leeuw, and van Wichelen 2014; Boersma and Schinkel 2015). Furthermore, recent research on transparency in the context of the IND has focused on the discretionary powers of civil servants making decisions about asylum requests (Severijns 2019), and on the "transparency" of the IND organization with regard to Dutch society and lawmakers (De Leeuw, Geerdink, and Smits 2019). Even the perspectives of volunteers working in Dutch asylum-seeker centers are heard in debates about the bureaucracy involved in processes of forced migration (Larruina and Ghorashi 2016). Yet the voice of the forced migrant seems to be largely missing in discourses on data practices in government bureaucracy.

Recent research that does include the perspective of the forced migrant usually deals with the period after permission to stay has either been granted or denied. It often details the difficulties of starting a new life in a foreign country (van Heelsum 2017; Huizinga and van Hoven 2018) or the precariousness of being denied citizenship (Kalir 2017; Boomgaard 2017). Instead, in this chapter, we intend to shed light on the mundane and often taken-for-granted nature of bureaucracy. While this part of the migration process might, from the perspective of the policy-maker, be understood as a moment in which people merely go through the necessary motions in order to collect "objective" information on the basis of which a decision can be made, for forced migrants these moments are very personal and stressful, and feel very uncertain, regardless of the intended transparency. Here, we aim to "let the subaltern speak" (Spivak 1988) and engage with the experiences of Syrian refugee women through metaphors such as "glass," "transparency," and "opacity." By highlighting not so much the intentions of the IND, but rather how various policies are turned into practice, we subscribe to social research interested in so-called "street level bureaucracies" (Lipsky 2010). More specifically, rather than investigating the experiences of various street-level bureaucrats, such as police officers (see for example Çankaya 2012), judges (Prescott 2009), or asylum officers (Darrow 2015; Dahlvik 2018), our research is aligned with studies that take the experiences of various marginalized identities, such as welfare recipients (Hansen, Lundberg, and Syltevik 2018), LGBT people (Nisar 2020), or refugees (Bhatia 2020), as the starting point for an investigation of a particular bureaucratic infrastructure. By letting the subjects speak rather than the data itself, we adhere to a feminist and postcolonial ethics of care, and aim to relate data practices of the IND to "human meaning-making, context-specificity, dependencies and temptations, as well as benefits and harm" (Leurs 2017, 150). In doing so, we will highlight how the collection of information by IND is anything but mundane, and instead a scary, and opaque process that offers very little opportunity for asylum seekers to express themselves in ways they desire.

The basis of this mixed-methods study of the information and transparency practices of the IND will be Araa's own IND file, her autoethnographic work, and her ethnographic work with four other Syrian refugee women. This means that this chapter will account for multiple voices and be narrated at times in the first person, namely with Araa speaking directly to account for her autoethnography, the voices of the four other Syrian women interviewed, as well as the scholarly contribution of Sandra Ponzanesi and Gerwin van Schie, co-authors of this piece. To remain narratively close to Araa's actual experiences and conversations, we will consistently write these parts from the first-person perspective, similar to the vignette at the start of this introduction. The personal IND file of each of the individual women functioned as the starting point for the interviews. These files consist of the following types of data: 1) a print-out of several database checks that the IND routinely does to find out if people are registered as criminals in the Netherlands or the EU, or have previously requested asylum in another EU state; 2) the transcripts of the interviews with the IND; and 3) the documentary evidence submitted by the asylum seeker, such as passports, birth certificates, newspaper clippings, and other documents that prove they have been Syrian citizens and were in danger of persecution.

Situated Knowledges and Autoethnography

This chapter combines the expertise of scholars working on critical data studies, gender, and migration studies in order to position the research done on Syrian refugee women in their interaction with the IND system within a postcolonial ethics of care, through which we allow the voices of the respondents to emerge in their own right. Yet the positions of mediator, interpreter, and facilitator that Araa fulfilled with, for, and among the respondents exemplify the important function of what Spivak has defined as the "native informant." In ethnography, Spivak points out, the native informant "is a blank" deprived of autobiography, but who enables the inscription of the Other by the West. The major problem with the history of post-Enlightenment theory lies in its own autobiography: it has used the figure of the native informant to mask "subjective structures" as "objective truth." Second, in doing so, it has uncritically assumed as unproblematic the subjectivity of the Other who consolidates the knowing Western subject and provides him or her with indigenous information. Spivak's observations also show that behind the figure of the native informant lie the questions of knowledge, power, and representation, the questions that still dominate current discussions on how to theorize and empower the Other without falling into the pitfalls of essentialism and binary opposition (Spivak 1999). For that reason, we will refer to feminist standpoint theory through which we take women's lived experiences, particularly experiences of caring and work, as the starting point for scientific inquiry. Feminist standpoint theorists, drawing on the work of scholars such as Donna Haraway, Sandra Harding, and Patricia Hill Collins, make three principal claims: (1) knowledge is socially situated; (2) marginalized groups are socially situated in ways that make it more possible for them to be aware of things and ask questions than it is for the non-marginalized; (3) research, particularly that focused on power relations, should begin with the lives of the marginalized (Harding 2008, 2009; Haraway 1991; Hill Collins 1997; Bracke and Bellacasa 2007).

Central to all these standpoint theories are feminist analyses and critiques of relations between material experience, power, and epistemology, and of the effects of power relations on the production of knowledge. Haraway's notion of situated knowledge is related, but also slightly different from, that of standpoint theory. It problematizes both subject and object, but instead of granting a privileged position to those at the margins, the subjugated knowers, it prefers to attribute privilege to partiality. This epistemological shift underscores the fact that "situated knowledge" is more dynamic and hybrid than other epistemologies that gives full knowability to the position of the subject. These situated knowledges involve "mobile positioning" (Haraway 1991, 192). Subjectivity is instead performed in and through the materiality of knowledge and practice of many kinds (Butler 1990, 1–34). Haraway says that situated knowledges require thinking of the world in terms of the "apparatus of bodily production." The world cannot be reduced to a mere resource if subject and object are deeply interconnected. Bodies as objects of knowledge in the world should be thought of as "material-semiotic generative nodes," whose "boundaries materialize in social interaction" (Haraway 1991, 201). The move to grant agency to material objects places the epistemology of situated knowledges at the center of scholarship in science and technology studies (Callon 1986; Latour 1987).

In light of this approach, it was not necessary for the researchers to have access to the personal information of the research subjects involved with this study as they themselves were in control of what they wanted to share. The only file that we, the co-authors of this chapter, looked into was Araa's file. In what follows we will critically examine the data practices of the IND's asylum procedure and juxtapose these practices with the stories of Araa and four other Syrian refugee women who have been through this process in order to problematize the concept of transparency and highlight its gendered implications. Therefore, we foreground an ethnographic approach through which the experiences of the Syrian women emerge, in alliance and conflict with their memories, "everyday geographies of belonging" (Huizinga and van Hoven 2018; van Liempt and Staring 2020), "politics of location" (Kirby 2015) and "media strategies" (Udwan, Leurs, and Alencar 2020). As Araa said herself, she felt very alienated by the IND process and this has motivated her need to analyze and put into context her experience along with that of many other women like her in order to make sense of the process. We therefore propose to analyze below the ethnographic method we used, along with the critical approach to data studies, which jointly reveal the opacity and shortcomings of the immigration system.

Analyzing the women's files and paying attention to their experiences during their investigation gives us an insight into how the algorithmic system has been used to process their asylum files. At the time the women were not well or sufficiently informed about the procedure and the implications of their responses. This put them in a subaltern position, based on misunderstanding and weaker communication skills, often mediated by an imperfect interpreter.

> I didn't know that I had the right to request an interpreter or officer of a specific gender, and I did not suffer from the gender of the interpreter, but I suffered from the interpreter's lack of knowledge of politics and the Syrian dialect, so I asked them to change the interpreter more than once because I was aware that the interpreter does not know the difference between the Syrian Communist Party and the Communist Party and the Communist Labor Party, so I was hearing the officer's questions that were different from what I expected or they were confusing, so I had to explain the differences between the parties and go through some detail to explain to the interpreter, and that was a new task I had to take on, which is to explain things to the interpreter because his information and translation are not sufficient, as he is not an expert on Syrian political issues. [Samira]

By analyzing the data that is shared and combining it with the stories of Syrian women who are in or have been through the IND's decision-making process, we produce situated knowledges (Haraway 1988; Harding 1991) centered on the experiences of the women as data subjects. The project has focused on Syrian *women* because research has shown significant differences in the experiences of women and men in the process of awaiting a decision (Nolin 2017). This is partly due to the continuation of power inequalities between genders in the countries of origin, which means that women often take care of the children and/or are expected to follow the commands of the men. Another issue is the vulnerable position of women staying in refugee centers (Spijkerboer 2017). By focusing on the experiences and stories of the women themselves, this chapter aims to emphasize the agency of individual research subjects in deciding what to share and what not to share, both with the IND and with the researcher.

The stories included in this chapter show the complicated process of asylum applications, marked by many circumstances that these women take with themselves from their home, during their journey, and upon their arrival in the Netherlands. The five women include Araa, already introduced above and the co-author of this chapter. She is 45 years old, married and has three children. She holds a PhD from the University of Damascus, Syria, obtained in 2012. In 2013, because of the war in Syria, she fled to the Netherlands and got her status permit. She was the founder and director for many years of the Syrian Women Foundation SVNL, which has allowed her to be of support to many Syrian women newly arriving in the Netherlands or struggling to find their balance in the new society. This also allowed her to gain access to the other women interviewed and win their trust.

The other women are: Laila Abdel, Lana Mustafa, Samira Sayed, and Salma Ezz El-Din.[1] Laila Abdel is a 48-year-old Syrian refugee. She is a graduate of the Faculty of Physical Education at the University of Damascus. She is a divorced woman who has been living with her daughter in the Netherlands for five years. She has two other children in Syria who live with their father. Lana Mustafa, 27 years of age, is a Syrian refugee who has been living in The Netherlands for five years. She has studied at the economics institute in Syria and came to the Netherlands to study computer sciences. She is married and has one child. She came to the Netherlands with her two-year-old daughter, and then used the right (family reunification), which allows her to bring her husband to the country of asylum. Samira Sayed is a 63-year-old Syrian refugee, who has lived in the Netherlands for six years. She is married with one son, who is now 38 years old. She has a BA in Arabic from the University of Aleppo in Syria. Salma Ezz El-Din is a 35-year-old Syrian/Palestinian refugee. She is a graduate of the Institute of Cooking and Nutrition, married with four children. She has been living in the Netherlands for four and a half years. All the women respondents have a good level of education, are highly literate and were competent in navigating bureaucratic and administrative hurdles back in their home country. Nonetheless, their class and educational background did not equip them for the intricacy and opacity of the Dutch immigration system and datafication process. We can expect, therefore, that this disadvantage and power imbalance would only be heightened and magnified for women with a lower educational background.

Entering the System: Becoming an Asylum Seeker

> When I went through the smuggling experience and got to know smugglers working between Turkey and Europe, I was offered several options for being smuggled. I chose the most expensive price, because I was afraid of sexual blackmail and other forms of extortion. I travelled to the Netherlands hidden in a truck, experiencing all the stages any refugee has to pass through to reach Europe. The moment you stand up in front of the asylum reception center to define yourself as a refugee coming from a country at war is still a vivid memory. Here, I experienced the moment that I turned from a citizen belonging to a country to a refugee in a new country. The moment I registered in the Dutch asylum system as a refugee, I experienced the Cartesian dualism of the ontological division between the "I" as a body, and the "I" as data subject. I felt strange in the Dutch system that understood me as a criminal or dangerous newcomer. The feeling that you are totally disregarded as a human being generated a reaction in me that motivated me to

1 These names are all pseudonyms at the request of the women concerned.

> do this research, and share my experiences and those of other women refugees who underwent the same procedure. [Araa Al Jaramani]

The moment the refugee women knocked on the door at the central asylum seekers' shelter (*Centraal Orgaan opvang Asielzoekers*, from now on COA) in Ter Apel, the Netherlands, and filed their request for asylum, many processes, both human and computational, were started. The first page of Araa's IND file lists several of the computational checks under the abbreviations OPS, HKS, N/SIS, BVV, Eurodac and Havank (see figures 1 and 2, summary in Table 1). These abbreviations do not mean anything to the uninformed, let alone newly arrived forced migrants fleeing from a war. They did not get any explanation or interpretation that would help them understand these acronyms and coding. These acronyms are not even explained in the margin of their investigation documents. While Samira and Salma did not notice the abbreviations on their file, Lana did not dare to ask the officer about it:

> Maybe if I were now under investigation I would have asked, but at that time I would not have dared to ask about its meaning. Knowing the meanings of these abbreviations will be better for me in that I will get to know the results or small details of what the investigators have discovered, as knowing that the abbreviations say that I am not involved in crimes in the European Union or in Syria during the war – I know that I am not guilty because I know myself, but it would have been good to know that the abbreviations on my papers mean that I am innocent. That they know who I am, instead of continuing investigations and me not being sure of what is happening around me, and always striving to absolve myself and prove that I am only a woman fleeing from war, who wants to protect her child from war. [Lana]

There was much apprehension about daring to ask questions, accompanied by the uncomfortable environment. Uninformed women feel embarrassed about their ignorance of these abbreviations and prefer not to ask. Samira described herself as a fool:

> I read the codes now and I see that I was a fool, and I see that there was no transparency, which is really annoying. I did not notice these symbols at the time so I did not ask about them. [Samira]

The framework of the IND data system relies on the Accountability Guidelines, which depend on the algorithmic system and which are applicable when asylum seekers start the process of applying for asylum status. The key questions are: first, whether the IND data system classifies Syrian refugee women according to the same cultural standards, irrespective of their gender or background, social, cultural and personal; secondly, whether the data system recognizes that refugees do not fully comprehend Dutch bureaucratic processes and that there is a gap in knowledge and understanding between Syrians and the Dutch system. This gap leads us to question not only the authority and correctness of the IND system that governs the interviews with these women, but also to wonder whether Syrian refugee women are entirely aware of their rights and power to object during the IND interviews.

Figure 1: First page of Araa's asylum request file

POLITIE

• Groningen

Postadres Postbus 107
9400AC ASSEN

Regio Groningen
Korpsonderdeel 1 Unit Vreemdelingenpolitie
Behandeld door [redacted]
Telefoon 088 - 167 16 55
Fax 088 - 167 16 29
V-nummer / Zaak [redacted] - 01/08/2013
Datum 01/08/2013
Onderwerp HV23 Controle personalia in systemen en registers

Controle personalia in systemen en registers

De personalia van hieronder genoemde vreemdelinge

Achternaam	: al Jaramani
Voorna(a)m(en)	: Ara
Geboortedatum	: 07/07/1975
Geboorteplaats	: Al-Mardj
Geboorteland	: Libië
Nationaliteit(en)	: Syrische
Geslacht	: vrouwelijk

werden door de korpschef van het regionaal politiekorps Groningen gecontroleerd in de volgende systemen/registers.

Systeem/register	**Datum/tijdstip controle**
N-SIS	01/08/2013 te 16:03 uur
OPS	01/08/2013 te 16:03 uur
HKS	01/08/2013 te 16:03 uur

Uitslag controle:

De personalia van betrokkene komen niet voor in N-SIS.

De personalia van betrokkene komen niet voor in OPS.

De personalia van betrokkene komen niet voor in HKS.

Bijzonderheden externe systemen:

Plaats: Ter Apel Datum: 01/08/2013

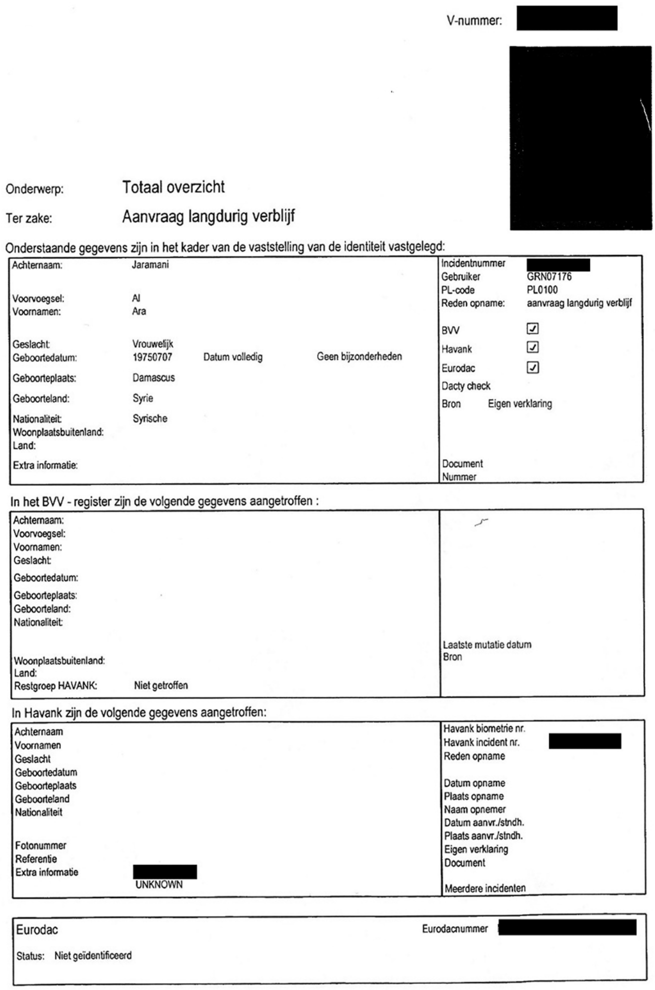
De korpschef van regionaal politiekorps Groningen,
namens deze,
de buitengewoon opsporingsambtenaar,
[redacted]
Noord Nederland
Unit Vreemdelingenpolitie

V.nr: [redacted] Pagina 1 van 2

[redacted]

Source: image from private documentation

Figure 2: Second page of Araa's asylum request file



V-nummer: [redacted]

[redacted]

Onderwerp: Totaal overzicht

Ter zake: Aanvraag langdurig verblijf

Onderstaande gegevens zijn in het kader van de vaststelling van de identiteit vastgelegd:

Achternaam:	Jaramani			Incidentnummer	[redacted]
Voorvoegsel:	Al			Gebruiker	GRN07176
Voornamen:	Ara			PL-code	PL0100
Geslacht:	Vrouwelijk			Reden opname:	aanvraag langdurig verblijf
Geboortedatum:	19750707	Datum volledig	Geen bijzonderheden	BVV	☑
Geboorteplaats:	Damascus			Havank	☑
Geboorteland:	Syrie			Eurodac	☑
Nationaliteit:	Syrische			Dacty check	
Woonplaatsbuitenland:				Bron	Eigen verklaring
Land:				Document	
Extra informatie:				Nummer	

In het BVV - register zijn de volgende gegevens aangetroffen :

Achternaam:		
Voorvoegsel:		
Voornamen:		
Geslacht:		
Geboortedatum:		
Geboorteplaats:		
Geboorteland:		
Nationaliteit:		
Woonplaatsbuitenland:		Laatste mutatie datum
Land:		Bron
Restgroep HAVANK:	Niet getroffen	

In Havank zijn de volgende gegevens aangetroffen:

Achternaam		Havank biometrie nr.	
Voornamen		Havank incident nr.	[redacted]
Geslacht		Reden opname	
Geboortedatum		Datum opname	
Geboorteplaats		Plaats opname	
Geboorteland		Naam opnemer	
Nationaliteit		Datum aanvr./stndh.	
		Plaats aanvr./stndh.	
Fotonummer		Eigen verklaring	
Referentie		Document	
Extra informatie	[redacted] UNKNOWN	Meerdere incidenten	

Eurodac	Eurodacnummer [redacted]
Status: Niet geïdentificeerd	

Source: image from private documentation

Table 1: List of different database checks performed as first step of an asylum request

System Abbrev.	Full name	Function
N-SIS	Schengen Information System	European system circulating "alerts about wanted or missing people and objects such as stolen vehicles and documents."[2]
OPS	Opsporingssysteem (Detection/Tracking System)	National Database of missing persons, missing vehicles, and crime suspects
HKS	Herkenningsdienstsysteem (Recognition Service system)	Outdated and controversial National Database of crime suspects and unsolved crimes[3]
BVV	Basisvoorziening Vreemdelingen (Basic Provision for Migrants)	Database of personal information of people who are involved in any process of migration to the Netherlands[4]
Havank	Het Automatisch Vinger Afdrukkensysteem Nederlandse Kollektie (Automatic Fingerprint System – Dutch Collection)	Dactyloscopic database of Dutch delinquents and former delinquents
Eurodac	European Dactyloscopy	European anonymous dactyloscopic database of people who have requested asylum in the European Union

As it turns out, refusing to cooperate in proving information can result in an immediate denial of the asylum request. One horrifying detail of the Dutch asylum law states that both "refusing to provide fingerprints" and "making the process of finger printing impossible by mutilating one's finger tips on purpose" may be considered grounds for the verdict of "insufficient cooperation" (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties 2021, translation by authors). Laws such as these show the large power imbalance playing out in day-to-day asylum procedures. While forced migrants routinely have to prove that they are in need of care and are, in fact, not criminals, the "right to remain silent" or any other right pertaining to the refusal to provide information does not apply. Even when the IND's need for information is in conflict with the personal beliefs or culture of a person requesting asylum, there is little room to deviate from protocol:

> I remember that while taking fingerprints, especially the photo of the face, there were difficulties for women because the woman had to reveal her ears and raise part of her hijab to make her forehead visible, and therefore pictures of the women

> were taken that did not respect their beliefs, and the women were terrified that they had to lift their veil, even partially. The women in the photos place were worried that their husbands and families might not believe they had to show their ears and part of their head in the pictures. [Lana]

Many of the IND regulations and criteria are not transparent and accessible to the refugees, who are at a disadvantage in understanding the implications of their choices, or consenting to the various steps in the procedure. Confusion about technicalities, coding, and acronyms used go together with the hazard of what gets lost in translation and the sensitive nature of intercultural communication that often leads to these women feeling intimidated, subalternized, and agreeing to choices and ticking boxes that they barely understand. But their desire and need to obtain asylum make them comply with any requests, even answering personal, invasive questions and showing compliance with a criminalizing line of questioning:

> The employee was scary, even if he was not wearing military clothing; he was looking clearly and intentionally into my eyes and observing the extent of my sincerity in the statements. And him opening my Facebook page and finding out that I was lying drove me to feel nervous so I was always saying that I am a refugee and not a suspect and that when I lied about the fact that I arrived in the Netherlands through Schiphol Airport, it was only to avoid separating me from my child who arrived 21 days before me and that according to Dublin law she is entitled to remain in Netherlands but I must stay in Belgium. I did not lie to hide a crime, but to ensure that I remain with my daughter. The employee's style of interrogating me and confirming the story of my escape from Syria made me feel that I was under investigation, not in an interview. [Laila]

There is, in addition, another crucial factor: the women often do not have knowledge or understanding of the new ways of automated computation that make people/refugees part of algorithmic systems that sort, classify, and process refugees – through biometrics and other systems of datafication. This makes them particularly vulnerable not only to the manipulation and breaching of their data (Madianou 2019) but also to having to consent to these procedures, for fear of rejection, without being fully informed, accountable, and responsible for the forms of datafication undertaken. In this sense there is not only a total lack of transparency dictated by the asymmetric relationship of power, but also a lack of technological expertise and awareness.

> I didn't know I had the right to reject it; perhaps I was ignorant of its importance, too. But now that I know the concept of privacy, and that I have the right to refuse having my information shared and that research parties may take some of my information and share it in a way that offends me, for example, now I know that if I am subject to such investigations, I will ask the investigator to explain to me

> who are these authorities and then I would search for the identity, credibility, and conformity of these transparency requirements and the reason why they want my personal information. [Samira]

The striking point is that during the investigation not one of our interviewees, including Araa, objected to their information being shared with other stakeholders, and no one asked about the nature of those stakeholders. On the contrary, when we read and discussed the IND investigation documents together, they emphasized many times how they assumed that they had no choice or say in the way the procedure was conducted. They assumed that they could not choose to refuse or object to the IND officer's request because this could lead to their status being rejected.

> If I look back to the period of my investigations by the IND, I am fully aware that I couldn't choose to refuse the IND's request to have my fingerprints or information, because at that time I thought that would affect my asylum application results. What people do not know is that the person who applies for an asylum application feels weak and that he/she cannot refuse any request from the authorities to which he/she has applied for asylum, for fear that his/her application would be rejected. [Lana]

However, the women managed also to find a form of agency and development during the procedure which allowed them to evaluate more carefully the options and opportunities ahead. We will discuss these in more detail in the next section.

Being Processed: The Interviews

Violation is one of the characteristic features of many asylum seekers' stories. That includes Laila and Lana, who both suffered violence and sexual abuse on their trip to Europe, and had to live in camps with strangers, while their restricted financial condition allowed smugglers to sexually blackmail them. They also have something else in common, which is their need to protect and defend their daughters as well as themselves. Though they are both asylum seekers now, Laila comes from a background as an activist, as she was a political prisoner under Assad's regime. Therefore, their memories about the violation they faced are different and this impacted their response to the IND investigator, who was not fully informed of the abuses they suffered before coming to the Netherlands. Laila told the investigators everything she felt when the Syrian intelligence agents forced her to be naked in front of them:

> They laughed and said that they want to examine my body. [Laila]

But she could not tell them the name of her smuggler, or even that he raped her and blackmailed her:

> The smuggler told me that the day had come, and he asked me to take with me only a simple bag, I met him in Omonoia Square. There he told me that he would take me to a small house while the plane was arriving, so I entered the house and he closed the door with the key, he took off my clothes, and raped me and I can still smell his dirty smell. He was strong and he was an Arab. I was afraid that if I told the investigator his name, I might be chased by the gang he belonged to and that they would harm me and my daughter, so I told the investigator that I do not remember his name. He is scary so I had no solution but to give him an extra sum to let me get out of the place. [Laila]

The decision of Laila to hide the identity of the smuggler during the IND investigation is an example of her mistrust of the IND. Moreover, she was intimidated by the IND officer, who accused her of lying about how she arrived in the Netherlands, not via Schiphol but through Brussels airport. Laila explained Araa her reason for lying: she was trying to rescue her daughter from the smugglers' harassment so she sent her to the Netherlands with a friend who had been residing there for the past two years, and she found her through a mutual friend as she was looking for a destination to flee to. After her daughter had arrived in the Netherlands, Laila did not find a way to get to the Netherlands but only to Brussels, where she later crossed the border into the Netherlands. According to the Dublin Conventions, the country of entry is the country where you must apply for asylum, which meant she would be separated from her daughter.

While the IND officer was complying with the Conventions of the European Union, with the implicit bias of protecting "Fortress Europe" and the Dublin Regulation from bogus migrants (Dijstelbloem and Meijer 2011; Leurs and Smets 2018), Laila struggled with understanding the mistrust and suspicions concerning her statements. It was difficult for her to be open-hearted, and convey her emotions, fears, and traumas. It was hard to explain that her situation was not only generated by her becoming a refugee all of a sudden, but also by the unfair gender system of her society, which considers her to be property and therefore not free in her choices and always to be judged by her behavior.

She could not explain why she did not wear a veil in the Netherlands when the IND officer doubted her identity because of a veiled photo on her passport. She was confused by this accusation, as if she had to explain that she was not an extremist coming to the Netherlands with the wrong intentions. She could not readily describe how her rejection of the veil was motivated by the rejection of her society that mistreated her as a woman. After the Syrian intelligence agents forced her to be naked, she took off the veil and refused to wear it again. But these are explanations that are difficult to bring forward during an IND hearing whose intimidating and accusatory tone does not help traumatized women to open up.

Likewise, Lana's testimony was shaded with insecurity and fear. She was shocked when, during the interview, the IND officer did not pay any attention to the abuse she encountered, either to the sexual violence she had experienced from the smuggler during her journey as a refugee, or even to the fact that she was raped by a refugee living in the same camp as her in Greece. But on the other hand, he was interested in knowing why she selected the Netherlands as a destination:

> In fact, there were questions that I considered strange and without reason. What was the reason for wanting to know why I came to Netherlands while I was waiting to be asked about the rape incident that I went through and I was waiting for them to protect me from the man who chased me after I moved from the camp and assaulted and violated me for the second time, I was waiting for them to arrest him. I did not receive help with anything and I do not know how these questions were beneficial for them. [Lana]

Despite the fact that Lana has been living in the Netherlands for five years, trying to empower herself by learning Dutch, doing karate courses, and continuing her education in ICT, she is still looking for some security to guarantee the honor and the safety of her and her daughter. At the same time, she does not care anymore about the veil in order to be respected by her society. She explained to Araa that she had changed her perspectives on life in particular after she had been raped twice and had encountered a Dutch bureaucratic structure that treated her with no special consideration given to her gender exploitation; she compared this to the attitude of her husband, who rejected her decision to be unveiled! And he is threatening to kidnap her daughter and take the girl back to Syria, because a woman without a veil does not deserve to raise her daughter.

When Lana was asked whether she thought that things would have been different if it was her husband who has applied for asylum first, she commented on the community context:

> For example, my husband and I study the Dutch language within the framework of the integration program in the Netherlands. We put our daughter in kindergarten during our school hours. Our families do not blame him for putting the child in kindergarten while he is out. Rather, they consider this a necessary matter that allows my husband to learn the language and to get a job, but they believe that I am not taking care of my child because I leave her upbringing to the kindergarten, and that I can learn the language online without having to leave the house and leave the girl in kindergarten. The same thing if he is the one who immigrated with the child, they will consider him a hero who migrated for the safety and future of his daughter, and whatever the results are, they will always grant him the honor of

> trying. But I did not find that in their conversations with me: I did not hear a single word of flattery, but always blame, intimidation, and recommendations. [Lana]

Many women may not have the necessary financial resources to undertake the flight journey to Europe. And as "Fortress Europe" continues to raise its barriers, it is more and more likely that asylum seekers will need to enlist the aid of traffickers or smugglers to help them enter Europe, and the high cost of this may well be beyond many women's reach. It can be argued that all these obstacles mean that women only leave their homes and families when circumstances become so hostile that they cannot possibly remain (Spijkerboer 2017).

Being "Integrated": Becoming a Data Subject

In 2013 Araa's asylum request was granted, and while this was a great relief, the transition from being under investigation to a position with agency and legal rights does not always directly translate into advantages in lived experience. When the General Data Protection Regulation (GDPR) was implemented by the European Commission in 2018, Araa and Gerwin decided that it would be interesting to test the implementation of this law at the IND. One of the provisions of the GDPR, aimed at increasing transparency and accountability, is that all people, as data subjects, have the right to request a copy of all information an organization or company keeps about that person.[5] This has led many governmental organizations, including the IND, to 1) design a clear formal route for making these requests on their website, as well as 2) appoint a central person responsible for the handling of data and privacy-related matters, often referred to as a "Data Protection Officer" (DPO). In August 2018, out of sheer curiosity, Araa and Gerwin filed a GDPR request, following the procedure explained on the IND website, sending an information request and a picture of Araa's passport to verify her identity.[6] However, the IND GDPR procedure proved to be much less straightforward than promised on the website of the IND.

The first obvious hurdles are the two very specific routes IND offers to file an information request, namely an online form that is only accessible in Dutch and a regular letter via mail (not email). The former option might not be helpful for many asylum seekers as the majority of them only start to learn this language when they arrive in the Netherlands. In many cases similar to Araa's, the Netherlands is not even the chosen destination, but rather a coincidental end point of their trip. While the latter option, an old-fashioned letter, does provide an opportunity to send a

5 See https://gdpr.eu/what-is-gdpr/.

6 See https://ind.nl/en/Pages/privacy.aspx.

request in a different language, it still presupposes people being able to navigate the Dutch mail system, something that, again, would be much easier for a Dutch-speaking person. Understandably, competency levels between asylum seekers vary greatly, since the opportunities to learn Dutch are heavily dependent on the personal situation in terms of financial means, level of education, and competency levels in other languages such as English. In addition, as shown by the story of Lana in the previous section, opportunities to learn Dutch can also heavily depend on gendered expectations concerning the necessity to learn a new language. By making competency in Dutch a requirement for filing an information request, the IND effectively also gendered the possibility for transparency and accountability.

The second issue is the lack of trust in government agencies that many refugees have due to experiences with corrupt and totalitarian systems. They know many stories of situations in which rights on paper did not mean anything in practice. This was no different when Gerwin first suggested to Araa that she should file a GDPR request with the IND:

> In Syria, I had written a drama text about corruption in certain Syrian universities. The university that I was doing my Ph.D. at wanted to punish me and tried to prevent me from getting to discuss my thesis, while I had already gotten that permission in the past. This put me in a challenging position, where I had to convince my university that Work is Work, and Education is a Right, and that they aren't allowed to punish me like that just due to my work and that I had every right to discuss my thesis. At the time I didn't discuss all of that with Gerwin, because I felt like I had to make this transition from a Syrian citizen into a Dutch citizen and I decided to go forward with the GDPR request. However, when I was on my way home, on the train, I wondered if I was wrong, and had put myself and my family at risk again. [Araa]

In hindsight, this situation shows the large difference in government trust between Gerwin and Araa. Where Gerwin did not consider the information request and Araa's refugee status to be related matters, Araa was afraid her request would be seen by the IND as a nuisance at best and a reason for denial of Dutch citizenship at worst. Instances like these should therefore be carefully considered by both researchers and organizations in their interactions with forced migrants. Even if researchers are right in their assessment of an information request being an entirely separate matter from bureaucratic decision-making processes, the precarious situation of an asylum seeker might cause unnecessary anxiety and stress. Depending on the aim of an information request, such negative effects are often not worth the potential benefits.

A final issue concerning the information request of Araa was that her interaction with the IND concerning her information did not end after she had sent the request. At first, IND did not respond at all and it turned out that the GDPR compliance promised by the IND website did not directly translate into action by the organization. After three months of waiting (more than a month past the legal waiting time for a response), Araa sent the IND a reminder via email. Instead of a response by email, Araa received a phone call from the IND Data Protection Officer. At first Araa was frightened by his direct approach, which was in sharp contrast to the distant bureaucratic process she had expected. However, she was quickly comforted when he immediately apologized and invited Araa to come over for coffee at his office in the building of the Department of Justice in The Hague:

> It was comforting to hear the employee's apologies and see him trying to find a fair solution for both of us. His kindness was a striking thing for me, especially when he invited me to have coffee at his office, gave me a tour of the IND building and took time to discuss the research that I wanted to work on at Utrecht University. [Araa]

After this meeting, Araa received a PDF file of about 150 pages which turned out to be an identical copy of all the combined documents Araa had already received during her asylum procedure. Until this day it is unclear to us in what ways Araa's invitation related to her work at Utrecht University at the time. In addition, we are aware of Araa's privileges in terms of her language proficiency in both English and Dutch, as well as her connections with the two other authors of this chapter, as we do not think that every forgotten GDPR request will end in an invitation for coffee and a tour of the IND building.

Conclusions

In this chapter we have analyzed the ways in which Syrian refugee women respond to the bureaucratic system of the IND, and how they tackle the increasingly datafied society and system of governance that impact them upon their arrival in the Netherlands. The aim was not only to analyze the structures, procedures, and decision-making process of the IND system but also to provide an account of the ways in which it is experienced and understood by Syrian refugee women themselves, with a particular focus on gender, ethnicity, class, and language issues. The IND is often referred to as an opaque, aggressive, and bureaucratic system that is far removed from the emotional and personal needs of the refugees.

There is a big discrepancy between the claim of transparency, objectivity, and fairness promoted by the IND as a governmental agency and the sense of opaqueness, and an impersonal and indifferent approach as experienced by the refugee women themselves. The Syrian women refugees feel like they have to defend themselves and prove their authenticity and trustworthiness, and show that they are deserving of help in order to gain access to Europe. This puts them in a constant position of being subject to scrutiny and suspicion. Due to algorithmic and machine-learning registration systems, the migrants often become dehumanized and reduced to statistical data. By providing an insight into the lived experiences, trajectories, and immigration stories that accompany this small but representative group of Syrian refugee women, we have attempted to offer a sobering and multiperspectival account of the limits and pitfalls of datafication systems.

We have done so by offering a theoretical framework of how datafication works in the case of the IND and the immigration system, and how a response can be formulated by drawing from standpoint theory that foregrounds the role of native informants and the ways in which partial knowledge can be produced that offers a more insightful and ethical response to the skewed relations imposed by the governmental system. Drawing on autoethnography and combining it with traditional ethnographic work, the chapter offers an exclusive insight into the procedures of the IND as experienced by the "data subject" through the different stages, from entering the system and asking for asylum, to being processed through the method of the interview, to finally becoming integrated and assimilated into the system when fulfilling all required criteria.

The chapter makes an important contribution to critical data studies, gender studies and in particular situated knowledges and ethnographic methods, legal and migration studies. It shows the relevance of the IND immigration procedure from the perspective of the data subjects. It sheds insights into possible ways of bridging cultural differences and increasing understanding and mutual trust under conditions of vulnerability and precariousness. We have tried to discuss how regulations and classifications around the garnering of refugee status from the point of view of the IND need to be counteracted by bringing forward and taking into account the perspective of the women who underwent the process themselves. The various accounts show the difficulties of aspiring citizens who struggle with linguistic, cultural, and legal barriers while dealing with emotional and traumatic experiences. We claim that it is important to consider the voices, opinions, and coping mechanisms of migrant refugees in order to envision new possible strategies for integration and assimilation based on mutual understanding of and respect for international agreements, but also intercultural practices.

To conclude, the chapter presents a moving as well as very informative collection of responses, experiences, and insights from five Syrian women refugee women who are in, or have been through the IND's decision-making process, and who speak back to the system, producing alternative knowledges and representations to the dominant and mainstream stories of migration and integration in the Netherlands.

References

Bakker, Linda, Sin Yi Cheung, and Jenny Phillimore. 2016. "The Asylum-Integration Paradox: Comparing Asylum Support Systems and Refugee Integration in The Netherlands and the UK." *International Migration* 54 (4): 118–32. Accessed April 15, 2021. https://doi.org/10.1111/imig.12251.

Bhatia, Monish. 2020. "The Permission to Be Cruel: Street-Level Bureaucrats and Harms Against People Seeking Asylum." *Critical Criminology* 28 (2): 277–92. Accessed April 15, 2021. https://doi.org/10.1007/s10612-020-09515-3.

Boersma, Sanne, and Willem Schinkel. 2015. "Imagining Society: Logics of Visualization in Images of Immigrant Integration." *Environment and Planning D: Society and Space* 33 (6): 1043–62. Accessed April 15, 2021. https://doi.org/10.1177/0263775815598153.

Boomgaard, Moniek. 2017. "Precarious Waiting: Time, Gender and 'Illegality' for Denied Asylum Seekers in the Netherlands." Master Thesis, Budapest: Central European University.

Bracke, Sarah, and Puig de la Bellacasa. 2007. "The Arena of Knowledge: Antigone and Feminist Standpoint Theory." In *Doing Gender in Media, Art and Culture,* edited by Rosemarie Buikema and Iris van der Tuin, 39–53. London and New York: Routledge.

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. London and New York: Routledge.

Callon, Michel. 1986. "The Sociology of an Actor-Network: The Case of the Electric Vehicle." In *Mapping the Dynamics of Science and Technology*, edited by Michel Callon, John Law, and Arie Rip, 19–34. London: Palgrave MacMillan.

Çankaya, Sinan. 2012. *De controle van marsmannetjes en ander schorriemorrie. Het beslissingsprocess tijdens proactief politiewerk*. Amsterdam: Boom Lemma.

Castles, Stephen, Hein de Haas, and Mark J. Miller. 2013. *The Age of Migration: International Population Movements in the Modern World*. Houndsmills, Basingstoke: Palgrave Macmillan.

Dahlvik, Julia. 2018. *Inside Asylum Bureaucracy: Organizing Refugee Status Determination in Austria*. Springer Open. Accessed April 15, 2021. https://doi.org/10.1007/978-3-319-63306-0.

Darrow, Jessica H. 2015. "Getting Refugees to Work: A Street-Level Perspective of Refugee Resettlement Policy." *Refugee Survey Quarterly* 34 (2): 78–106. Accessed April 15, 2021. https://doi.org/10.1093/rsq/hdv002.

De Leeuw, J.F., J.H. Geerdink, and L.J.E. Smits. 2019. "Onderzoek Vermoeden van Misstand Bij de IND." The Hague: Onderzoekscommissie vermoeden misstand bij de IND.

De Leeuw, Marc, and Sonja van Wichelen. 2014. "Doing 'Integration' in Europe: Postcolonial Frictions in the Making of Citizenship." In *Gender, Globalization, and*

Violence: Postcolonial Conflict Zones, edited by Sandra Ponzanesi, 145–60. London and New York: Routledge.
Dijck, José van. 2014. "Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology." *Surveillance & Society* 12 (2): 197–208.
Dijstelbloem, Huub, and Albert Meijer, eds. 2011. *Migration and the New Technological Borders of Europe*: *Migration, Minorities and Citizenship*. Houndsmills, Basingstoke: Palgrave Macmillan.
Douglas, Scott, and Albert Meijer. 2016. "Transparency and Public Value: Analyzing the Transparency Practices and Value Creation of Public Utilities." *International Journal of Public Administration* 39 (12): 940–51.
Es, Karin van, and Mirko Tobias Schäfer. 2017. "New Brave World." In *The Datafied Society: Studying Culture Through Data*, edited by Mirko Tobias Schäfer and Karin van Es, 13–22. Amsterdam: Amsterdam University Press.
Gillespie, Marie, Souad Osseiran, and Margie Cheesman. 2018. "Syrian Refugees and the Digital Passage to Europe: Smartphone Infrastructures and Affordances." *Social Media + Society* 4 (1): 1–12. Accessed April 15, 2021. https://doi.org/10.1177/2056305118764440.
Haar, Marleen van der, and Dvora Yanow. 2011. "Allochtoon Als Metafoor En Categorie. Over de Handelingsimplicaties van Beleidstaal." *Beleid En Maatschappij* 38: 160–78.
Hansen, Hans Krause, and Mikkel Flyverbom. 2015. "The Politics of Transparency and the Calibration of Knowledge in the Digital Age." *Organization* 22 (6): 872–89. Accessed April 15, 2021. https://doi.org/10.1177/1350508414522315.
Hansen, Hans-Tore, Kjetil Lundberg, and Liv Johanne Syltevik. 2018. "Digitalization, Street-Level Bureaucracy and Welfare Users' Experiences." *Social Policy & Administration* 52 (1): 67–90. Accessed April 15, 2021. https://doi.org/10.1111/spol.12283.
Haraway, Donna. 1991. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In *Simians, Cyborgs and Women: The Reinvention of Nature*, edited by Donna Haraway, 149–81. New York: Routledge.
Harding, Sandra. 2008. *Sciences from Below: Feminisms, Postcolonialities and Modernities*. Durham and London: Duke University Press.
Heelsum, Anja van. 2017. "Aspirations and Frustrations: Experiences of Recent Refugees in the Netherlands." *Ethnic and Racial Studies* 40 (13): 2137–50. Accessed April 15, 2021. https://doi.org/10.1080/01419870.2017.1343486.
Hill Collins, Patricia. 1997. "Comment on the Hekman Truth and Method: Feminist Standpoint Theory Revisited: Where is the Power?" *Signs: Journal of Women in Culture and Society* 22 (2): 375–381.
Huizinga, Rik, and Bettina van Hoven. 2018. "Everyday Geographies of Belonging: Syrian Refugee Experiences in the Northern Netherlands." *Geoforum 96* (2018):

309–317. Accessed April 15, 2021. https://doi.org/10.1016/j.geoforum.2018.09.002

Kalir, Barak. 2017. "State Desertion and 'out-of-Procedure' Asylum Seekers in the Netherlands." *Focaal* (77) 1: 63–75. Accessed May 3 , 2021. https://www.berghahnjournals.com/view/journals/focaal/2017/77/fcl770106.xml?rskey=Ex5kU9&result=1.

Kirby, Andrew. 2015. *The Politics of Location: An Introduction*. Routledge.

Larruina, Robert, and Halleh Ghorashi. 2016. "The Normality and Materiality of the Dominant Discourse: Voluntary Work Inside a Dutch Asylum Seeker Center." *Journal of Immigrant & Refugee Studies* 14 (2): 220–37. Accessed April 15, 2021. https://doi.org/10.1080/15562948.2015.1131877.

Lathrop, Daniel, and Laurel Ruma, eds. 2010. *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly Media.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society.* Cambridge, MA: Harvard University Press.

Leurs, Koen, and Kevin Smets. 2018. "Five Questions for Digital Migration Studies: Learning From Digital Connectivity and Forced Migration In(to) Europe." *Social Media + Society* 4 (1): 1–16. Accessed April 15, 2021. https://doi.org/10.1177/2056305118764425.

Leurs, Koen, 2017. "Feminist Data Studies: Using Digital Methods for Ethical, Reflexive and Situated Socio-Cultural Research."*Feminist Review*. 15 (1): 130–154.

Van Liempt, Ilse, and Richard Staring. 2020. "Homemaking Within and Beyond the Places Assigned to Syrian Refugees in the Netherlands." *Geographical Review* 111 (2): 308–326. Accessed April 15, 2021. https://doi.org/10.1080/00167428.2020.1827935.

Lipsky, Michael. 2010. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. New York: Russell Sage Foundation.

Madianou, Mirca. 2019. "Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises." *Social Media and Society* 5 (3): 1–13.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin Harcourt Publishing Company.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. 2021. "Vreemdelingencirculaire 2000 (C)." *Circulair*. Accessed April 15, 2021. https://wetten.overheid.nl/BWBR0012288/2021-01-14.

Nisar, Muhammad Azfar. 2020. "Phenomenology of the Stop: Street-Level Bureaucracy and Everyday Citizenship of Marginalized Groups." *International Review of Administrative Sciences* 86 (2): 316–32. Accessed April 15, 2021. https://doi.org/10.1177/0020852318776363.

Nolin, Catherine. 2017. *Transnational Ruptures: Gender and Forced Migration*. London and New York: Routledge. Accessed April 15, 2021. https://doi.org/10.4324/9781315235714.

Prescott, Dana E. 2009. "Unified Family Courts and the Modern Judiciary as a Street-Level Bureaucracy: To What End for the Mythical Role of Judges in a Democracy." *QLR* 27 (1): 55–112.

Rath, Jan. 2001. "Research on Immigrant Ethnic Minorities in the Netherlands." In *The Politics of Social Science Research: 'Race', Ethnicity and Social Change*, edited by P. Ratcliffe. Houndmills, Basingstoke and Hampshire: Palgrave Macmillan. Accessed April 15, 2021. https://doi.org/10.1057/9780230504950_7.

Sánchez-Querubín, Natalia, and Richard Rogers. 2018. "Connected Routes: Migration Studies with Digital Devices and Platforms." *Social Media + Society* 4 (1): 1–13. Accessed April 15, 2021. https://doi.org/10.1177/2056305118764427.

Schinkel, Willem. 2013. "The Imagination of 'Society' in Measurements of Immigrant Integration." *Ethnic and Racial Studies* 36 (7): 1142–61. Accessed April 15, 2021. https://doi.org/10.1080/01419870.2013.783709.

Severijns, Ralph Willem Jan. 2019. *Zoeken naar zekerheid: Een onderzoek naar de vaststelling van feiten door hoor- en beslismedewerkers van de Immigratie- en Naturalisatiedienst in de Nederlandse asielprocedure*. Staat en Recht (SteR). Deventer: Wolters Kluwer.

Spijkerboer, Thomas. 2017. *Gender and Refugee Status*. London and New York: Routledge.

Spivak, Gayatri Chakravorty. 1988. "Can the Subaltern Speak?" In *Marxism and the Interpretation of Culture*, edited by Cory Nelson and Lawrence Grossberg, 271–313. Urbana and Chicago: University of Illinois Press.

Spivak, Gayatri Chakravorty. 1999. *A Critique of Postcolonial Reason: Toward a History of the Vanishing Present*. Cambridge, MA: Harvard University Press.

Ticktin, Miriam. 2006. "Policing and Humanitarianism in France: Immigration and the Turn to Law as State of Exception." *International Journal of Postcolonial Studies*, 7 (3): 346–368.

Udwan, G., Leurs, K., & Alencar, A. 2020. "Digital resilience of Syrian refugees in the Netherlands. Social Media for Social Support, Health and Identity." *Social Media + Society*. 6 (2): 1–11. Accessed April 15, 2021. https://doi.org/10.1177%2F2056305120915587.

III: Digital Care

Everyday Curation? Attending to Data, Records and Record Keeping in the Practices of Self-Monitoring[1]

Kate Weiner, Catherine Will, Flis Henwood, and Rosalind Williams

"Data Power" and the Turn to Everyday Monitoring

The growth in apps, wearables and networked technologies that measure or keep track of a plethora of bodily states, actions, and experiences, has been referenced in a number of key discussions within social sciences. Self-monitoring has been characterized as disciplining and normalizing, creating particular kinds of neoliberal, self-regulating subjects and reinforcing obligations for self-care (e.g., Lupton 2016). For some, it is seen as part of the broader "datafication of health" (Mayer-Schoenberger and Cuckier 2013; van Dijck and Poell 2016; Ruckenstein and Dow Schull 2017), in which, increasingly, aspects of bodily experience are transformed into quantified data. Self-tracking data may be seen as "lively" (Lupton 2016, 2018a) as they are aggregated, analyzed, circulated and potentially repurposed. Scholars, for example, have drawn attention to the commodification of these data (e.g., van Dijck and Poell 2016; Ajana 2018) and their potential contribution to surveillance, allowing, for example, health professionals access to individuals' conduct (Lupton 2012). The terms dataveillance and lateral surveillance are also used in this context, signaling the more diffuse network of actors amongst whom data may be circulated, including individuals who may willingly share their data with their own social networks (Andrejevic 2005; Rich and Miah 2017).[2]

The foregoing scholarship has been characterized as being centrally concerned with "data power" (Kennedy 2018). Offering some critique of this, Ruckenstein and Dow-Schull (2017, 256) call for more attention to everyday engagements with data in practice: "Scholars who attend to the power dynamics of datafication have been

1 This chapter was first published as Weiner, Kate, Catherine Will, Flis Henwood, and Rosalind Williams. "Everyday Curation? Attending to Data, Records and Record Keeping in the Practices of Self-Monitoring." Big Data & Society (volume 7, issue 1, January 2020) under Creative Commons license CC-BY-4.0.

2 For a fuller account of these literatures, see Ruckenstein and Schüll (2017).

faulted for their heavy focus on the oppressive, normalizing, and exploitative forces of datafication and their lack of attention to cases of noncompliance, appropriation and existential possibility."

Kennedy (2018, 20) similarly argues that discussions of datafication tend to leave "little scope for agentic engagements with data." One response has been to turn to more ethnographically informed research. Gaining an understanding of everyday or mundane engagements with self-monitoring and the data that emerge, it is suggested, is important to inform both scholarship on, and policy and commercial expectations about, the role of data in society (Pink et al 2017; Weiner et al. 2017; Kennedy 2018; Gorm and Shklovski 2019).

There is now a blossoming scholarship on everyday or mundane self-monitoring, often addressing fitness, exercise, or food tracking, but also other areas including self-monitoring of chronic health conditions. A number of related themes are emerging in this scholarship and here we draw attention to three in particular. The first takes seriously people's emotional engagements with self-monitoring data (Ruckenstein 2014; Pantzar and Ruckenstein 2015; Lupton 2017), countering images of those who self-monitor as impartial, rational actors pursuing health aims (see Pantzar and Ruckenstein 2015; Lupton 2016, 2017). This has included discussion of the enjoyment or pleasure derived from self-tracking, associated with for example seeing personal successes or supporting a self-identity as a fit or healthy person, as well as disappointments, worry or frustration when these are not achieved (e.g., Whitson 2013; Ancker et al. 2015; Lomborg and Frandsen 2016; Pink et al. 2017; Urban 2017; Lomborg et al. 2018; Lupton 2018a, 2018b, 2019; Gorm and Schlovski 2019).

A second theme concerns the different values attributed to data derived from self-tracking. In some instances, value is seen to derive from the (normalized) knowledge claims it allows, for example in the ability to detect patterns, or lend credibility to facts (Fiore-Gartland and Neff 2015). However, data have also been shown to have communicative value, as a way to connect with others, or share intimate stories (Fiore-Gartland and Neff 2015; Sharon and Zandbergen 2017). In other instances, self-tracking may be linked to mindfulness and awareness of one's own body and experience. Here, the act of monitoring or recording may be as, if not more important than, reviewing aggregated data (Nafus and Sherman 2014; Sharon and Zandbergen 2017). Scholars have also drawn attention to the situated and embodied way people make sense of, or assess the value of, tracking data in relation to other ways of knowing, as well as the way emotions are intertwined with these valuations (Nafus and Sherman 2014; Lupton 2018a; Lupton et al. 2018).

A third theme relates to the hidden or invisible work (Star and Strauss 1999) of making data and allowing it to travel. Pink et al. (2018), for example, are interested in the often obscured or hidden work of mundane repair. Introducing the idea of "broken data" and "repair work," they argue for ethnographic attention to the constitution of digital data, describing a process of improvisation or repair to fill in the inevitable gaps in people's self-tracking data. For example, people use multiple devices or use devices in unexpected ways such as using a step counter to record cycling. In this way, they suggest a focus on making sure data are coherent for oneself with no responsibility to provide accurate data to each device or app (Pink et al. 2018). In her work on a digitized, algorithmic physical rehabilitation system, Schwennesen (2019) also enrolls "repair work"' to describe the way patients tinker with the system to make it work in practice. Other scholars draw attention to the broader work of engaging in self-monitoring, beyond generating data[3], that remains invisible to its proponents (Ancker et al. 2015; Lupton 2018b, 2019).

This discussion of emotional engagements with and different values of data, and the work of making data, go some way to restoring a degree of agency to those who self-monitor. It helps to complicate narratives about the disciplining and normalizing power of self-monitoring practices and about the flows of self-monitoring data related to the potential for surveillance and/or commodification. The ethnographically-informed work, in the tradition of user studies (Oudshoorn and Pinch 2003), therefore provides empirical research "from below" that helps to nuance the "data power" argument. At the same time, some of this work also considers the agency of things/devices, which we discuss in more detail below.

In this chapter we aim to extend important work concerning everyday data practices and, specifically, the everyday constitution of digital data (Pink et al. 2017 2018; Lupton 2018a). Taking inspiration from and building on the concepts of "broken data" and "repair work" (Pink et al 2018), we adopt and develop the idea of *curation* in relation to self-monitoring, using material from our study of the everyday practices of monitoring blood pressure and/or Body Mass Index (BMI).

Adding a Curatorial Lens

Curation is multivalent. Davis (2017) offers a theoretical treatment of digital curation, describing curation as a theory of attention, concerned with how people allocate and control attention. Drawing on examples relating to social media, she suggests that curation "broadly…refers to the discriminate selection of materials for display [online]," where "productive curation" involves deciding what to "document, make, share, and with whom" and is integral to performances of self for oneself and for others (Davis 2017, 771, 772).

While there has not been a thoroughgoing application of the notion of curation to self-monitoring, a sense of this selectivity is present in some emerging

3 Such as learning and routinizing techniques, or making sense of and assessing the accuracy of the data

studies of everyday self-monitoring practices. Kent (2018, 67), in a study of how self-trackers represent "health" through social media, discusses the way her participants construct an appropriate self-tracking persona "through careful inclusion and exclusion of certain health information." Studies of calorie and of fitness tracking have documented the way participants may manipulate data input, for example not recording everything consumed on days of excess (Didžiokaitė et al. 2018; Lomborg et al. 2018), or not saving "unflattering" runs (Esmonde 2020, 84). They might also engage in "episodic use" (Gorm and Shklovski 2019) of tracking technologies, recording calories or wearing fitness trackers only on days when they anticipate good, interesting or useful numbers (Didžiokaitė et al. 2018; Lupton et al. 2018; Esmonde 2020; Gorm and Shklovski 2019). In this way participants are selective in the records they create either imagining an external audience, or supporting their motivation and protecting themselves from disappointing outcomes.

In Nielsen's (2015) work, the external audience is particularly important for patients, who she suggests undertake "filtration work" when making entries to a new e-health system. This involves being selective in relation to what information to provide and has a particular, dialogic, orientation; patients imagined the receiver, and shaped their entries in line with conversations they hoped to pursue or avoid. Work on the development of a clinical self-monitoring system for diabetes similarly showed how patients might decline to share data or respond to clinicians' messages (Piras and Miele 2017). All of the studies discussed so far illustrate selectivity in records made or shared, suggesting there is value in the concept of curation in relation to self-monitoring.

In considering the value of curation as a conceptual lens in this context, we need to acknowledge that curatorial work is suffused with and inseparable from the emotions associated with tracking and the value of the data. In our discussion above we have illustrated how people may gain pleasure or satisfaction and are able to communicate particular stories about themselves through the (hidden) work of curating their records. In this context, curation helps to bring together the three emerging themes we identified relating to self-monitoring, linking the hidden work of making data with the emotional aspects and the value of the data.

How does curation relate to the notions of repair work (Pink et al. 2018) and filtration work (Nielsen 2015)? All these concepts help to bring to light the hidden work of making data. While curation signals the possibility of selectivity, repair work is suggestive of an ultimate hope of completeness. Yet it does involve putting materials together for one's own satisfaction. Where curation may be broadly communicative, part of identity construction for oneself and conveying this to others, filtration work in Nielsen's (2015) account is solely orientated to others. It is concerned with opening up or closing down particular conversations with specific actors. In this paper, we would like to propose curation as an overarching concept, where repair work and filtration work offer particular examples of this concept. Curation helps to illuminate the hidden or underarticulated work of producing and sharing self-monitoring records. It, thus, helps to bring the agency of those who monitor into view. Yet the concept of curation does not only illuminate the work of human actors, but may also acknowledge the work of materials.

Curation as Socio-Material Practice

In her discussion of curation, Davis (2017, 775) attends to the agency of materials through making a distinction between "human" and "machine" curation, discussing how the design of platforms and algorithms shapes and constrains the way users produce and consume online content. Pink et al. (2017, 3) make a similar move in relation to self-monitoring data, wanting to take "users'" perspectives seriously but also to "decentre the human," suggesting that "personal data" are "constituted and experienced between human and digital/algorithmic devices and processes." The breakages in data they describe, when devices are not charged or lack connection, when software updates make existing devices redundant, or devices track some activities but not others, draw attention to the way devices and platforms shape the production of self-monitoring records. As we have discussed, their work also documents the way users may attempt to get around material constraints through their repair work (Pink et al. 2018). In previous research we have drawn attention to the multi-user functionality of some devices for measuring blood pressure and weight, to highlight the way these shape, or script (Akrich 1992) particular ways of recording and sharing data (Williams et al. 2020a). These sorts of socio-material analyses illuminate the way platforms and devices shape the production and management of self-monitoring records without resorting to technological determinism. They allow space for both users and technologies (and their developers) to have agency (Oudshoorn and Pinch 2003; Lupton 2018a; Henwood and Marent 2019).

Yet, in considering the material dimensions of curation we would like to draw attention to the kinds of self-monitoring so far discussed in critical scholarship. This has, with notable exceptions (e.g., O'Riordan 2017; Lupton and Smith 2018), tended to focus on digital and networked types of self-tracking involving, especially, fitness and diet apps and wearables. Yet, as Neff and Nafus (2016, 98) note, "self-tracking tools do not have to be fancy" and might include low tech materials such as pen and paper. Indeed, Fox and Duggan's (2013) oft-cited research reported that the majority of Americans who track a "health indicator" did this with pen and paper or "in the head." Rather than equating self-tracking with digital and networked self-tracking, we think it is important to consider the wider materials and technologies and their place in the practices of self-monitoring. What, for example, are the implications of different materials for data flows? Further, since curation is concerned with allocating and controlling attention (Davis 2017), are there material dimensions to paying attention, avoiding noticing or being inattentive to self-monitoring?

In sum, we propose that a curatorial lens facilitates the exploration of the way self-monitoring data are constituted in practice, illuminating the work of both humans and materials. Further, the idea of curation helps to link the work of making data with the emotional aspects of self-monitoring and the value of the data. In our analysis we adopt this lens to develop a socio-materialist account (Weiner and Will 2018; Henwood and Marent 2019; Williams et al 2020a, b) of everyday data practices relating to self-monitoring, exploring what records people keep, what materials are involved and whether and how records are shared. We suggest that this curatorial approach helps to clarify the relationship between self-monitoring and the accrual and flow of data. By paying attention to which data are or are not recorded, as well as the ways data are recorded, the research provides specificity to the ways in which self-monitoring may or may not contribute to Big Data sets in different ways. It allows reflection on the "'liveliness" (Lupton, 2018a) of self-monitoring data, in terms of their potential to be circulated, reconfigured and monetized, and do so in ways that might act back on the individuals who generated them. Ultimately, we propose curation can therefore be helpful in interrogating concerns with data power.

Methods

The chapter is based on a UK study involving interviews with people who self-identified as monitoring their blood pressure or BMI/weight. Our engagement with self-monitoring stemmed from our broader interest in everyday health practices, the use of health technologies in domestic settings and the way these might redistribute health work between the home and the clinic (see Weiner et al. 2017; Weiner and Will 2018; Henwood and Marent 2019; Williams et al. 2020a, b). Home blood pressure monitoring and BMI monitoring offer particularly interesting cases in the way they blur the boundary between the clinic and the home.

In the UK there are well established consumer markets for both blood pressure and BMI monitoring. A range of devices are available to purchase in supermarkets, pharmacies, and online retailers, such as digital blood pressure monitors, digital and analogue weighing scales and digital body analysis scales. These products include both stand-alone and networked devices and may be accompanied by proprietary apps, but also paper booklets or diaries for recording readings (see Williams et al. 2020a, b, for further analysis of this market). There are also other apps to calculate/track BMI or track blood pressure, such as MyFitnessPal and Apple Health, where data may be entered manually or pushed through from networked devices, as well as websites providing online BMI calculators. Both forms of monitoring have clear links to clinical interests. Monitoring blood pressure is well established in clinical practice and self-monitoring is increasingly sanctioned as one response to white coat hypertension (doctor-induced high blood pressure) (NICE 2011). Clinical concern with BMI and weight relate to obesity, and to risks of diabetes and cancer and forms part of public health messages (Gatineau et al. 2014; Hooper et al. 2016). In sum, both have clear clinical relevance and established self-monitoring markets.

While our study involves voluntaristic self-monitoring, we acknowledge the non-innocence of self-monitoring technologies and their links with broader socio-political contexts. Notwithstanding the contested history of BMI, the measure links with weight management which is associated with strong narratives of personal responsibility, guilt and shame (Lupton 2013). This and other forms of tracking intended to work on the body relate to gendered norms of beauty and fitness as well as to health (Esmonde 2020). Relatedly, there are clinical/psychological concerns about the possible links between food-tracking apps, such as MyFitnessPal, and eating disorders (Lupton 2018b). Discourses relating to tracking are also infused with assumptions about people's capacity to incorporate tracking which do not chime with gendered, classed, or marginalized experiences of daily life or work routines (Ancker et al. 2015; Esmonde and Jette 2020; Lupton 2018b). At the same time, there are concerns that fitness tracking may be pushed or imposed (Lupton 2016) by healthcare insurers or employers (Lupton 2016; Ajana 2018; Esmonde and Jetter 2020). We note, however, the relevance of these concerns is limited in the UK context, where healthcare is largely accessed through a universal, national, government-funded system. Even so, self-tracking is likely to be linked with uneven and differentiated experiences and effects.

In our study we made efforts to recruit a diverse sample. Following institutional ethics approval, we advertised on email lists at three UK universities and noticeboards across campuses, at older people's groups and at community centers in less-advantaged areas. The advert sought people who identified themselves as "measuring and keeping track" of either their blood pressure or BMI. In this paper we draw on 67 interviews conducted with 81 people, including 14 interviews with couples. Participants varied in terms of age, sexuality, ethnicity, socio-economic background, and health. All had acquired monitoring devices for themselves and no one reported acquiring these from employers or clinicians. While we were alive to issues of diversity, we did not find these significant in the current analysis, although they are more central to other themes (See Will et al. 2020).

In interviews, we asked people how they came to monitor or acquire a device, what they do or do not do with it and who else might use it, how this may have changed over time, and with whom data are shared. The limitations of "conventional" social science methods such as interviews for researching everyday life are well rehearsed (Martens, Halkier and Pink 2014, 3). People may find it difficult or are unable to talk about certain elements of their everyday practices, in particular embodied, tacit and affective aspects (Martens and Scott 2004; Martens et al. 2014). The use of material objects or photos in interviews can provide an aid to memory and reflexivity that interviews alone cannot elicit (e.g., Harper 2002; Woodward 2016). In our interviews we invited participants to demonstrate their monitors and talk through any records they kept and where these were stored. This helped both to prompt reflection and tie practices to particular time periods and events. We analyzed the interviews thematically (Hamersley and Atkinson 1995), collaboratively developing a coding frame, which synthesizes our theoretical grounding with emergent themes.

In this paper we focus on self-monitoring data practices and the materials this involves. It is not our intention to provide a definitive definition of self-monitoring and we do not see an obvious difference between this and self-tracking. Resonating with other research (e.g., Lupton 2019) we followed an emic approach, keeping our recruitment material broad and allowing people to identify themselves as engaging in self-tracking in order to study what this involves for them. Lupton (2016, 2) proposes that self-tracking entails "'practices in which people knowingly and purposively collect information about themselves which they review and consider applying to the conduct of their lives." In our analysis we explore the distinctions and relationships between these different potential aspects of self-monitoring focusing on three main themes: the relationship between taking and recording measures; how and where records are made; and storing and reviewing records.

Findings

The Relationship Between Taking and Recording Measures

No Records

We start our analysis by considering the approximately one-quarter of our participants who took measures but did not record these. Understanding curation to be concerned with attention, we consider what people are attending to in these cases. In other words, if they are not recording their data, what are they doing when they self-monitor? Sometimes participants did this for reassurance, just wanting to know if their blood pressure or BMI was in the normal range and they were able to recall this without needing to remember the precise number or to keep records. People talked of monitoring "to keep an eye on something" or "for peace of mind," illustrating the emotional resonance of the practice. For example, Gary explains he has anxiety issues and uses his blood pressure monitor for reassurance:

> I need to know if there's something wrong you know... So if I think I've got a bit of a headache or I get some palpitations I'll check it. (Gary, 45, school administrative officer, white British)

Gary does not record his readings, and cannot recall the precise numbers from the last time he used his monitor (4 weeks before the interview), but knows they were "under the 140 and 90" which he called "the bench mark."

For other participants, monitoring was concerned with managing day-to-day conduct. Linda, for example, does not keep a record of her weight and uses BMI as trigger to take action when she sees herself "creeping" near to the boundary between normal and overweight:

> If I see myself creeping...I haven't actually got to the point of going into the next category ... so it's sort of time to take some action in the sense of, you know, just cutting back on what I'm eating, being more careful about portion sizes, that sort of thing. (Linda, 67, retired Further Education teacher, mixed heritage)

Occasionally, blood pressure measures also resulted in immediate action such as drinking some green tea or trying to relax.

In these cases, people were not seeking to understand patterns in their data, but to attend to their immediate bodily status for reassurance or potential actions. Self-monitoring helps to address questions such as: how am I today? Am I stressed? Do I need to go to the doctors? Should I eat less today? This helps to explain why some participants cannot recall precisely or do not record or review their data.

Discerning Work and Partial Data

Yet among those who did record data, participants described selective recording, including not recording particular readings. Ayo weighs herself on stand-alone digital scales, and records this into her Samsung Health app, which calculates her BMI. She told us she only records her weight when it had gone down:

> Ayo: When I weigh and it's more and I've put on weight I don't enter into the app to update my BMI... I only do it when I lose weight
> Interviewer: So how come you don't put it in?
> Ayo: Because it makes me sad... The fact that I've put on weight, which is not what I want. I want to lose weight... So that's sad for me so I don't bother putting it onto... each time I found out I've lost weight then I add my weight here... I want to see that I'm losing weight on my app. (Ayo, 33, university researcher, black African)

Ayo's account again underscores the affective resonance of self-monitoring records, which have the capacity to make her "sad" if they go in the wrong direction. This chimes with studies of calorie and fitness tracking (Didžiokaitė et al 2018; Esmonde 2020; Gorm and Shklovski, 2019).

Other participants reported, in a more pragmatic vein, that there was no point noting down weight or blood pressure when it was stable, only noting when there is change. Annie records her blood pressure on scraps of paper and in a booklet. She told us she only notes this down when her reading is high:

> I don't write down the good numbers, I only write down the bad numbers. So when it's fine I don't bother, but when it's bad I think I probably should, because I've got a rubbish memory I think I probably need to keep a record of that. (Annie, 45, university administrative officer, white British)

So in contrast to Ayo, who only records her measures when they go in the right direction, Annie only records "bad" numbers. Participants measuring blood pressure also discussed processes of selecting or averaging multiple readings for recording e.g., best of three. We also encountered occasional stories of participants curating charts to make them more meaningful or pleasing, for example by removing outlying data points or choosing the span of time and the right axis. Gareth, for example, showed a graph of his weight on his Google Fit app, illustrating how the axis changes when he selects different years, and how the falling line pleases him:

> That's me overall graph. I'm quite pleased with how that's steadily falling. Especially when I put that year in it puts a different axis on it and it's whoosh. (Gareth, 58, property maintenance engineer, white British)

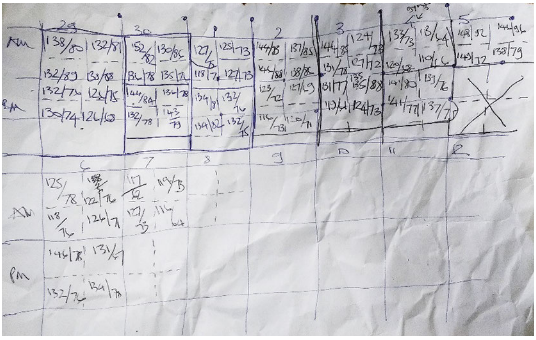
The foregoing accounts illustrate how participants are selective in the way they compile data into records. We propose that they are undertaking *discerning work* (rather than "repair work"), making judgements about which readings are useful, or worth remembering or drawing attention to, and how to process or clean readings to make best sense of them. Further, rather than the metaphor of "broken data," we propose the idea of *partial data* may be more apposite in these instances. Partial here has a double meaning, understood in the sense that only some of the data get recorded, but also in the sense of interested or partisan, in contrast to impartial or neutral. The readings written down or entered into apps may only be a subset of the readings taken and may be selected for very particular reasons.

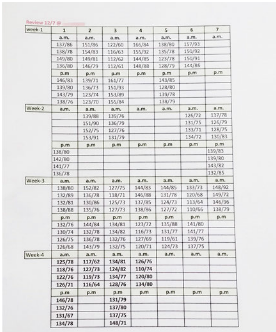
Intermittent Measures and Partial Data

In other interviews, participants told us of intermittent measurement which led to intermittent records. In interviews relating to blood pressure, participants sometimes compiled records for time-limited periods specifically to take to clinical consultations. For example, Fred records his blood pressure for one month prior to his appointment compiling a spreadsheet which he prints to take to his doctor's appointment. Such intermittent records resulting from intermittent measuring might be understood as a second form of partial data, in the sense that it is compiled from time to time[4].

Fred's account is useful in reinforcing an important point of our analysis by demonstrating the considerable work of making records. Fred told us that he records his blood pressure on pieces of paper by hand, transcribing these data into a spreadsheet which he compiles specifically for his appointment. He labels the spreadsheet with the name of his GP practice and the date of his appointment. This helps to illustrate the curatorial work involved in making data ready to share. In moving from hand-written slips of paper to a neatly presented spreadsheet, Fred demonstrates the skillfulness and probity of his self-monitoring. The materiality of the spreadsheet communicates "I am a responsible patient, you can take my readings seriously"[5].

Figure 1: Fred's handwritten notes, retrieved from the waste paper bin during the interview
Figure 2: Fred's spreadsheet compiled for his doctor's appointment, practice name blanked out





So far we have shown that the practices of taking a reading do not always lead to keeping a record of those data. For some of our participants, self-monitoring did not involve making any records. Where people do keep records, the readings entered into apps or recorded elsewhere may only be a subset of the readings

4 Intermittent measurement aligns with Gorm and Shklovski's idea of "episodic use," although the timeframes differ in the studies. Where "episodic use" denotes on and off use across days in the week, intermittent measurement in our study denotes periods of tracking and not tracking across months or years.

5 In further analysis we intend to consider whether we can see filtration work (Nielsen 2015) in relation to the kinds of conversations interviewees were hoping to pursue with their clinicians and the degree to which these data flow. We do not have space here to do justice to such analysis.

taken. We have suggested the term *discerning work* to describe the skillful judgements people make about which data to record, which to omit, and how to process and present records that differentially draw attention to successes, warning signs, or the credibility of the person making the record. We have also suggested that participants may create *partial data* both in the sense that they may choose only to record some of the readings they take, selecting these through the discerning work we have described, or in the sense that they measure, and therefore record, intermittently. From a user perspective, records may not be intended to be comprehensive or continuous, so here there is no "broken data" and therefore no need for "repair work." This means that, beyond the data breakages identified by Pink et al. (2018), there are other reasons why data may not flow seamlessly from measurement to an individual's records to be aggregated by third parties. All of this should act as a gentle corrective to expectations about the exploitation of such data in existing literature (see Ruckenstein and Dow Schull 2017; van Dijck and Poell 2016).

How and Where to Record

The analysis so far has mostly concerned which measures people record, but has also touched on the importance of how records are made and presented, particularly in our discussion of Fred. We now turn to this theme in more detail, to expand on issues around how and where people record and the materials involved. We will consider the socio-material arrangements in this curatorial work that draw attention to or deflect attention from different aspects of self-monitoring.

Visibility and Being Reminded

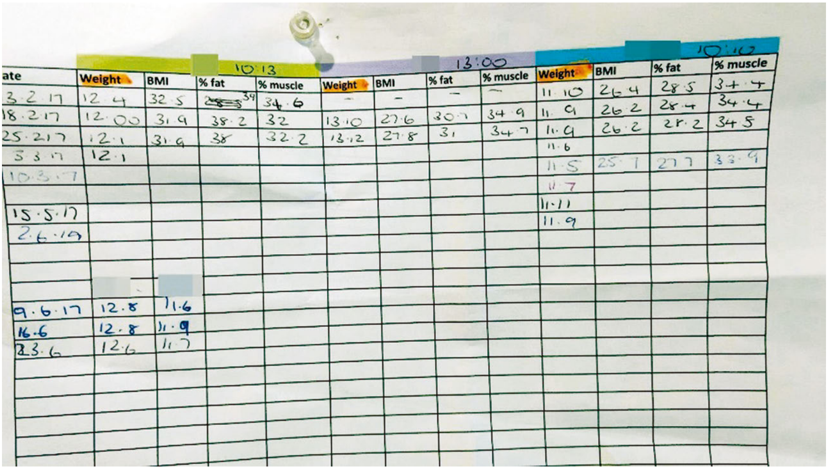
While half of the participants in the study had experience of using an app to track BMI, the visibility of paper records emerged as an important theme. For example, Becky told us she was losing weight together with her sister and sister's wife. They met on Saturday mornings to record their weight and kept a joint record on a sheet of paper. The record had been set up on a spreadsheet, but this was printed off and weights written on by hand. When asked why they did not simply enter the data onto the spreadsheet, Becky responded:

> I think because it was going to be a group thing that we could all jot it down while we were together. So I think that's why I have a physical sort of... (Becky, 36, charity researcher, white British)

In contrast to the digital co-presence discussed by Pink and Fors (2017), where people who are physically separate share data and are present together online, self-monitoring in our study involved physical co-presence where different materials come to the fore. A paper record in this account appears to allow these three women to participate together and to attend to their data collectively.

It was striking that records for sharing with partners, relatives and friends tended to be paper, charts and/or DIY forms of digitally networked communication such as texting or setting up WhatsApp groups. We encountered very little discussion of sharing through broader social media or proprietary self-monitoring apps that would facilitate sharing with wider social networks or publics, even where participants had devices with the capacity to do so. The materials our participants discussed appeared to allow them to do things together with limited, selected others (friends, family), and offer each other encouragement, whilst precluding broader attention. This finding resonates with some studies of digital self-tracking (Pink and Fors 2017, Lomborg et al 2018, cf. Kent 2018), placing into question expectations of widespread lateral surveillance (Rich and Miah 2017). Here we extend the existing analysis by considering not only with whom participants share, but also the materiality of sharing.

Figure 3: Becky's "group thing," names blanked out



The visible emplacement of monitoring devices in particular domestic spaces, for example close at hand on a table next to a favorite armchair, may encourage people to monitor (Weiner and Will 2018). In the same way, the emplacement of self-monitoring records, such as a chart pinned to a wall or a record on a mobile phone that is always to hand, might act as a reminder in different ways. Participants told us that leaving paper records and charts somewhere visible within the home helped to remind them to monitor or helped to keep commitments in mind. Becky, for example, told us that the shared record she made with her sister and sister-in-law was pinned to her notice board in her home office. She explained that this was visible enough to remind herself she was trying to lose weight, but not so public that visitors to the house would readily see it (compared with, for example, pinning it to the fridge in the kitchen). It is placed to hold her attention whilst avoiding drawing the attention of visitors. This concern with the emplacement of the chart suggests that even paper records have the capacity to act as "indiscreet technologies" (see Oudshoorn 2012), making public aspects of identity or practice that people would rather keep to themselves. In Becky's case, the materiality and emplacement of the record lend themselves both to monitoring in a group and to keeping a project in mind, whilst allowing the record to remain relatively private.

Others told us that they recorded on the phone because it is always with them, unlikely to be forgotten or because it formed a convenient mode for transporting records. Bella, for example, related that it was not until she got her smartphone and found a free blood pressure app that she started to record her readings. Before that she had not made a record:

> because it was a pain in the neck…Because I had to keep writing it down and then remember to write it down and find the piece of paper, like that. So I was really happy when I found the app. (Bella, 57, charity administrator, white British)

In contrast to a "piece of paper," Bella's phone is always nearby. Yet recording on phones did not always involve using proprietary tracking apps, as participants also told us they used note apps or Google Sheets to record self-monitoring data. Samuel, for example, talked about recording his blood pressure readings in a note app, to take to his doctor's appointment:

> Interviewer: you said you'd put the records on your phone for a time. Was there any reason for that?
> Samuel: Only for ease of transport, I knew I'd have my phone with me. I haven't kept them as a record, it was just a way of transporting information to the surgery with me.
> Interviewer: Okay, so was it an app for the blood pressure?
> Samuel: No, it was just a note. (Samuel, 62, university counsellor, white British)

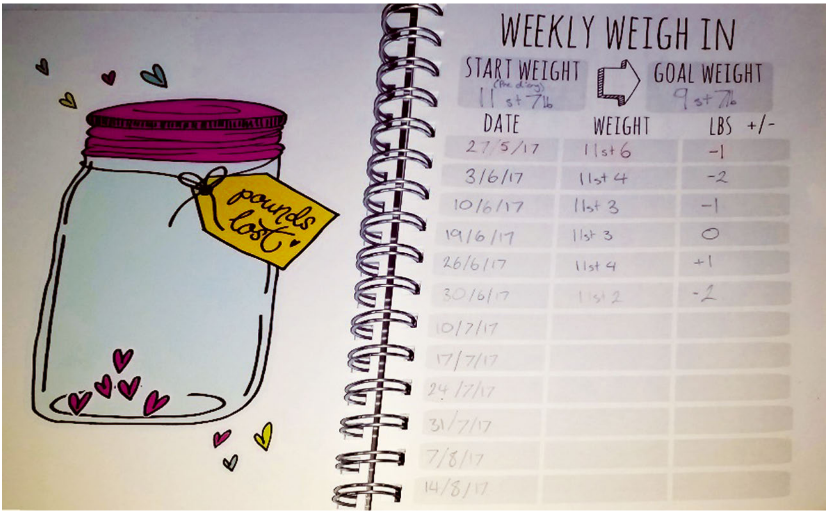
The constant presence of phones facilitated the recording of measures and made sure records were always at hand.

Not Being Reminded

While a few participants told us they had moved to apps from other ways of recording, in one instance a participant had moved away from an app, precisely because of the attention it demanded. Here again the emotional resonance of tracking comes into view. Andrea told us she used MyFitnessPal and that, while she continued to record her calorie intake in the app, she had stopped recording her weight in this because she found it contributed to her becoming "obsessive" about it. At that point she told us she preferred to record her weight on a weekly basis in a notebook:

> So although I weigh myself at the minute, I'm not putting into MyFitnessPal because I found I was getting maybe a bit too – I started weighing myself every day and I may have got a little bit too obsessive about it... I felt I've gradually started being calmer about it ... So I thought I'm going to gradually start doing all the things that I used to do again which is weigh myself once a week. (Andrea, 27, university administrator, white British)

Figure 4: Andrea's "weekly weigh in" note book



While MyFitnessPal is designed to allow and encourages daily recording of weight, Andrea's "weekly weigh in" notebook is quite literally scripted for weekly recording. In Andrea's case, she recounts being overly concerned with, or over-attentive to, her weight and in changing to a different way of recording tries to regulate this over-attentiveness.

In this section, we have paid particular attention to the materiality of records, looking at the different ways these contribute to paying and regulating attention to self-monitoring. Like others (Lomborg et al. 2018; Pink and Fors 2017), we found little sharing through proprietary self-monitoring apps, but the employment of other materials – paper, spreadsheets, WhatsApp group – which work to limit attention to small, selected groups of people. The visibility and emplacement of paper records may facilitate collective practice within the home and help to remind people to monitor or keep commitments in mind. The nearness of phones facilitates the recording of readings. Yet in one case, the immediacy of a mobile app was associated with over-attentiveness, and a paper record helped to remedy this. More baldly, the diverse materiality of self-monitoring records, including, but not limited to proprietary self-monitoring apps, suggests that self-monitoring data may not always be readily compiled or harvested by third parties, placing brakes on the liveliness (Lupton 2019) of this data.

Storing and Reviewing Records

Broken Data and Repair Work

Curation involves both what people make and what they keep and display. This relates to what they would like to remember or be reminded of. We found occasional stories of people going to efforts to retrieve data from different sources in order to retain a complete record. These may be understood as examples of the repair work and broken data proposed by Pink et al. (2018). John (55, IT support, white British) had recorded his weight and BMI on a weekly basis for the last decade using a number of different platforms. Initially he used a weight loss website called Weightloss Resource, because his wife was already subscribed to this. He ended this subscription in 2014, downloading his data to a Google Sheets spreadsheet, and moved to MyFitnessPal, which he used for 10 months before getting a Fitbit. He told us that to export his 10-months of data from MyFitnessPal would incur a fee, which he was not prepared to pay, although he lamented the "gap" in his data.

We asked him at different points in the interview why he had downloaded his data and if this was important to him. His responses suggested an emotional connection to graphs as "comforting." They also posed a link between records and biography – "a reminder of where you'd been and where you'd come to." John related this to one particularly significant time in his life, during which one of his daughters was diagnosed with and treated for a serious illness. For John, retaining and looking over his records appeared to be both a way to celebrate his successes in weight control and to remember how he and his family had come through his daughter's illness. This underscores the emotional and communicative aspect of these records.

Dormant Records

In considering self-monitoring through the lens of curation, we have so far discussed record keeping practices in fairly deliberate terms. We have portrayed self-monitoring records as being created and shaped through a combination of the discerning work of humans and the materiality of the devices and broader technologies involved. Our final brief section provides a caveat to this view, suggesting that sometimes the human and material elements combined in such a way that participants found it difficult to keep track of their records. For example, Tony keeps records of monitoring his blood pressure in a rather "haphazard" fashion on various slips of paper and backs of envelopes. He told us he stored most of his records in a bag but that he threw a lot of these away:

> Interviewer: Do you normally keep the readings in that?
> Tony: Yeah in a rather haphazard fashion. On bits of notes... I just used to write them down on bits of paper and shove them into this bag. And then I was packing this all away one day and suddenly thinking gosh there's an awful lot of ancient results here, I'm never going to do anything with these, and I remember I chucked a load away. (Tony, 54, electronic engineering lecturer, white British)

The emplacement of records "shoved" into a bag means that they do not seem to hold Tony's attention. Like Tony, participants were often not trying to discern patterns in their measures in any sustained way, nor did they look over them to derive comfort or pleasure. People talked of losing records and in some cases they rediscovered records during the course of the interview that they did not remember making or keeping. While these stories were most notable in relation to paper records, participants also talked of difficulties locating and retrieving digital records. Terry (83, retired credit controller, white British), for example, recounted that he plugs his digital blood pressure monitor into his computer every six months or so to look at the data, but that when he did this recently, prompted by receiving an invitation to participate in the study, he was unable to locate previous readings, telling us "I must have saved it somewhere, and I can't find it anywhere." He attributes this to having acquired a new computer.

One way to interpret these accounts of lost or inaccessible records is through the lens of broken data, characterized by ruptures in people's records. Yet, in these cases, these ruptures were not accompanied by efforts to repair the records except perhaps for the purposes of our interviews. Pieces of paper were stashed away in bags or with devices, computers and phones were upgraded, and old ones were discarded or moved to transitional spaces in the home such as the loft, just as old diaries were stored in the cellar. We find parallels within the sociology of consumption in Sophie Woodward's notion of "dormant things" (Woodward, 2015). This references the accumulation of things not currently being used which may be stored deliberately, but may also be forgotten. Drawing on Woodward, we propose these accumulated records might be considered *dormant*. While an important focus of our analysis has been to highlight people's agential engagements with the constitution of data, the notion of *dormant records* helps to acknowledge disengagements and lack of intentionality. Records may have been created and stored deliberately but become dormant when they no longer hold participants' attention.

Discussion

This chapter introduces and develops an analysis of self-monitoring through the lens of curation. In doing so it builds on and extends a now growing scholarship on everyday self-monitoring. By analyzing data practices through the concept of curation, we illuminate both the human work involved in making and retaining records, whilst, at the same time, taking seriously the role of materials. Understanding curation as a theory of attention, we have analyzed the different ways both humans and materials are implicated in drawing attention to, or detracting attention from, the practices of self-monitoring and the data these create.

In thinking through the work of curation, we have proposed the concepts discerning work and partial data in relation to self-monitoring. In suggesting these we have been influenced by Pink et al.'s (2018) concern with the way (digital) data are constituted in everyday situations. We find that their "concept metaphor" of broken data and focus on the "work of repair" do useful analytic work, although they describe only a small amount of the curatorial work we encountered in our study. The ideas seem to imply an aspiration for completeness which we find often absent. We think that discerning work in the context of self-monitoring provides a broader term for describing the work that people do to create self-monitoring records. We have shown how people do not necessarily record all the readings they take, but make decisions about which to record. In this way, records may be selective where people record only the data they are happy with, or that they feel they need to be reminded of. Here, data may be partial, but not necessarily broken, in the sense of representing an incomplete set of the data created and capturing the selectivity or interestedness of the data recorded. We have also suggested that data may be understood as partial when monitoring is undertaken intermittently, perhaps with specific purposes in mind (e.g., for a doctor's appointment) or in seemingly less patterned ways. We recognize that all data are partial (Gitelman and Jackson 2013), but think the notion of partial data, in contrast to broken data, helps to keep hold of this sense of selectivity and intermittency.

A second contribution of this chapter is our analysis of the material dimensions of curation in relation to self-monitoring. Rather than figuring self-monitoring as exclusively digital or networked, we have documented the variety of materials associated with records and pointed to the way different materials help to hold or regulate participants' attention. The visibility of paper records may facilitate people to monitor together when they are physically co-present. Notebooks or charts prominently emplaced might also help participants to remember to monitor or keep a commitment in mind. Self-tracking and other apps such as Google Sheets, and phone memos or notes helped to retain attention through their emplacement, always present and unlikely to be forgotten. Yet, as exemplified by one participant, the permanent presence of smartphones might risk people becoming over-involved in self-monitoring and the relatively static emplacement of paper records might enable monitoring to be kept at a distance. Further, when people shared self-monitoring records, these were mostly in the form of paper records and digital DIY networks. Compared with tracking apps, we suggest these are perhaps more straightforwardly discreet because sharing is more readily limited.

We propose that curation, as a theory of attention, helps brings together different aspects of self-monitoring discussed in the more ethnographically-informed scholarship. It links the work of making records (e.g., Pink et al. 2018) with the emotional aspects of self-monitoring (e.g., Ruckenstein 2014; Anckers et al. 2015; Pantzar and Ruckenstein 2015; Pink et al. 2017; Urban 2017; Lupton 2018a, 2018b, 2019; Lomborg et al. 2018, Gorm and Shklovski 2019) and what scholars have discussed as the different values of self-tracking data (Nafus and Sherman 2014; Fiore-Gartland and Neff 2015; Sharon and Zandbergen 2017; Lupton 2018a; Lupton et al.2018). In undertaking curation, people constitute records that are pleasing or communicate aspects of their identity or biography (e.g., a trustworthy patient, a successful dieter). Materials may help to distance self-tracking so as to reduce obsessiveness or anxiety, or may act as a reminder of a commitment. In this way, we have shown that curation complements other research on how people make sense of or evaluate tracking data (Lupton 2018a), by underscoring the way these valuations may prefigure and shape the generation of data in the first place.

While our analysis finds space for the agency of those who self-monitor in creating records, we have illustrated the difficulties some participants had marshalling unruly materials, as they decided what to keep or tried to remember if or where they had stored records. Following Woodward (2015), we have suggested the term dormant records to account for records that have been stored in case of potential future use, as well as those that are unattended and those that have been forgotten.

Our analysis has pointed to the way people engage and disengage with self-monitoring and the data that it produces. In this sense, it helps to put data and records in their place. Accounts of discerning work and partial data return a degree of agency to users of self-tracking technologies in the creation and circulation of data, while being attentive to the constraints imposed by the diverse materialities involved. Like others (Didžiokaitė et al. 2018; Lomborg et al. 2018; Esmonde 2020; Gorm and Shklovski 2019), we have shown that, even where people do record their data in ways that might be compiled by third parties (i.e. through apps), they do not necessarily give up all their data, and may be selective in what they record. They may also not be "hooked" (Lomborg et al. 2018) into continuous monitoring and recording (see also Gorm and Shklovski 2019), and therefore the data they produce may be limited, even if it is in a material form that can easily flow. The different materials enrolled for making and sharing records might further dampen expectations about the potentials for data flows.

To what degree does our analysis stem from cases we chose? Blood pressure monitoring and weight/BMI currently involve measuring devices that may be, but are often not, networked. Whether and how to keep records is relatively open. However, devices are likely to become increasingly networked or even wearable, suggesting a move from manual data input to system-generated records and more continuous measurement. Yet the discerning work in the creation of records we and others have observed suggests that people may still exercise a degree of agency over their self-tracking data. Further, as Lupton et al. (2018), Esmonde (2020) and Gorm and Shklovski (2019) amply illustrate in relation to activity tracking, people may still remove devices or delete unwanted data points, or only monitor on days that they think are likely to show desirable results. Moreover, as we have discussed, the materiality of records is entwined with the practices of monitoring. People's willingness to use specific technologies may depend on the levels of visibility and discretion they offer and the degrees to which they are suited to the types of individual or collective practices of monitoring we have described. This means that, even when they could use them, people may sometimes eschew digital/networked technologies and use analogue/non-networked forms of monitoring and recording instead.

What does this all mean for the generation of Big Data and our understandings of data power (Kennedy 2018)? Adopting a curatorial lens helps to unpack precisely which data points are recorded and omitted from self-monitoring records, and the ways in which these data may or may not travel beyond the people who generate them to be aggregated into Big Data sets and/or used by other actors. It thus adds specificity to discussions about data that does not become "big" and lends nuance to our understanding of the potentials for data flows in practices of self-monitoring. Acknowledging the importance of discerning work, the partiality of data, the varied materiality of self-monitoring and the dormancy of some records suggests we should temper expectations about data flows, data power and claims about surveillance and exploitation linked to these.

References

Ajana, Btihaj. 2018. "Communal Self-Tracking: Data Philanthropy, Solidarity and Privacy." In *Self-tracking*, edited by Btihaj Ajana, 125–141. Cham: Palgrave Macmillan.

Akrich, Madeleine. 1992. "The de-scription of technical objects." In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker and John Law, 205–224. Cambridge: MIT Press.

Ancker, Jessica S, Holly O Witteman, Baria Hafeez, Thierry Provencher, Mary van de Graaf, and Esther Wei. 2015. "You Get Reminded You're a Sick Person: Personal Data Tracking and Patients With Multiple Chronic Conditions." *Journal of Medical Internet Res*earch 17 (8). Accessed February 23, 2021. https://doi.org/10.2196/jmir.4209.

Andrejevic, Mark. 2005. "The work of watching one another: Lateral surveillance, risk and governance." *Surveillance & Society* 2 (4): 479–497.

Davis, Jenny L. 2017. "Curation: a theoretical treatment." *Communication & Society* 20 (5): 770–783. Accessed February 23, 2021. https://doi.org/10.1080/1369118X.2016.1203972.

Didžiokaitė, Gabija, Paula Saukko, and Christian Greiffenhagen. 2018. "Doing Calories: The Practices of Dieting Using Calorie Counting App MyFitnessPal." In *Metric Culture: Ontologies of Self-Tracking Practices*, edited by Btihaj Ajana, 137–155. London: Emerald.

Dijck, José van, and Thomas Poell. 2016. "Understanding the promises and premises of online health platforms." *Big Data & Society* 3 (1). Accessed February 23, 2021. https://doi.org/10.1177/2053951716654173.

Esmonde, Katelyn. 2020. "There's only so much data you can handle in your life: Accommodating and resisting self-surveillance in women's running and fitness tracking practices." *Qualitative Research in Sport, Exercise and Health* 12 (1): 76–90. Accessed February 23, 2021. https://doi.org/10.1080/2159676X.2019.1617188.

Esmonde, Katelyn, and Shannon Jette. 2020. "Assembling the 'Fitbit subject': A Foucauldian-sociomaterialist examination of social class, gender and self-surveillance on Fitbit community message boards." *Health* 24 (3): 299–314. Accessed February 23, 2021.https://doi.org/10.1177/1363459318800166.

Fiore-Gartland, Brittany, and Gina Neff. 2015. "Communication, Mediation, and the Expectations of Data: Data Valences Across Health and Wellness Communities." *International Journal of Communication* 9 (1): 1466–1484. Accessed February 23, 2021. https://ijoc.org/index.php/ijoc/article/view/2830.

Fox, Susannah, and Maeve Duggan. 2013. *Tracking for Health*. Pew Research Centre (Washington DC). Accessed February 23, 2021. https://www.pewresearch.org/internet/2013/01/28/tracking-for-health/

Gatineau, Mary, Caroline Hancock, Naomi Holman, Helen Outhwaite, Lorraine Oldridge , Anna Christie and Louisa Ells. 2014. *Adult Obesity and Type 2 Diabetes.* Public Health England (Oxford).

Gitelman, Lisa, and Virginia Jackson. 2013. "Introduction." In *Raw Data'is an Oxymoron*, edited by Lisa Gitelman, 1–14. Cambridge: MIT Press.

Gorm, Nanna, and Irina Shklovski. 2019. "Episodic use: Practices of care in self-tracking." *New Media & Society* 21 (11–12): 2505–2521. Accessed February 23, 2021. https://doi.org/10.1177/1461444819851239.

Hammersley, Martyn, and Paul Atkinson. 1995. *Ethnography: Principles in Practice*. London: Routledge.

Harper, Douglas. 2002. "Talking about pictures: A case for photo elicitation." *Visual Studies* 17 (1): 13–26. Accessed February 23, 2021. https://doi.org/10.1080/14725860220137345.

Henwood, Flis, and Benjamin Marent. 2019. "Understanding digital health: Productive tensions at the intersection of sociology of health and science and technology studies." *Sociology of Health & Illness* 41 (S1): 1–15. Accessed February 23, 2021. https://doi.org/10.1111/1467-9566.12898.

Hooper, Lucie, Annie Anderson, Alice Foster, Rosenberg Gillian, and Vohra Jyotsna. 2016. *Public Knowledge of the Link Between Obesity and Cancer.* Cancer Research UK (London).

Kennedy, Helen, 2018. "Living with data: aligning data studies and data activism through a focus on everyday experiences of datafication." *Krisis: Journal for Contemporary Philosophy.* 2018 (1): 18–30. Accessed February 23, 2021. http://eprints.whiterose.ac.uk/129959/

Kent, Rachael. 2018. "Social Media and Self-Tracking: Representing the 'Health Self'." In *Self-Tracking*, edited by Bithaj Ajana, 61–76. London: Palgrave Macmillan.

Lomborg, Stine, and Kirsten Frandsen. 2016. "Self-tracking as communication." *Information, Communication & Society* 19 (7): 1015–1027. Accessed February 23, 2021. https://doi.org/10.1080/1369118x.2015.1067710.

Lupton, Deborah. 2012. "M-health and health promotion: The digital cyborg and surveillance society." *Social Theory & Health* 10 (3): 229–244. Accessed February 23, 2021. https://doi.org/10.1057/sth.2012.6.

Lupton, Deborah. 2013. *Fat*. London: Routledge.

Lupton, Deborah. 2016. *The Quantified Self*. Cambridge: Polity.

Lupton, Deborah. 2017. "How does health feel? Towards research on the affective atmospheres of digital health." *Digital Health* 3 (1). Accessed February 23, 2021. https://doi.org/10.1177/2055207617701276.

Lupton, Deborah. 2018a. "How do data come to matter? Living and becoming with personal data." *Big Data & Society* 5 (2). Accessed February 23, 2021. https://doi.org/10.1177/2053951718786314.

Lupton, Deborah. 2018b. "I Just Want It to Be Done, Done, Done!: Food Tracking Apps, Affects, and Agential Capacities." *Multimodal Technologies and Interaction* 2 (2). Accessed February 23, 2021.https://www.mdpi.com/2414-4088/2/2/29.

Lupton, Deborah. 2019. "It's made me a lot more aware: A new materialist analysis of health self-tracking." *Media International Australia* 171 (1): 66–79. Accessed February 23, 2021. https://doi.org/10.1177/1329878x19844042.

Lupton, Deborah, Sarah Pink, Christine Heyes LaBond, and Shanti Sumartojo. 2018. "Digital Traces in Context| Personal Data Contexts, Data Sense and Self-Tracking Cycling." *International Journal of Communication* 12 (1): 647–665. Accessed February 23, 2021. https://ijoc.org/index.php/ijoc/article/view/5925/2258.

Lupton, Deborah, and Gavin J. D. Smith. 2018. "A Much Better Person: The Agential Capacities of Self-Tracking Practices." In *Metric Culture: Ontologies of Self-Tracking Practices*, edited by Btihaj Ajana, 57–75. London: Emerald.

Martens, Lydia, Bente Halkier, and Sarah Pink. 2014. "Researching habits: advances in linguistic and embodied research practice." *International Journal of Social Research Methodology* 17 (1): 1–9. Accessed February 23, 2021. https://doi.org/10.1080/13645579.2014.853999.

Martens, Lydia, and Sue Scott. 2004. *ESRC report – Domestic Kitchen Practices: Routines, Risks and Reflexivity* ESRC End of Award Report. Keele University (Keele).

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.

Nafus, Dawn, and Jamie Sherman. 2014. "This One Does Not Go Up To 11: The Quantified Self Movement as an Alternative Big Data Practice." *International Journal of Communication* 8 (1): 1784–1794. Accessed February 23, 2021. https://ijoc.org/index.php/ijoc/article/view/2170.

National Institute of Health and Care Excellence (NICE). 2011. *Hypertension in Adults. Diagnosis and Management. Clinical Guideline 127*. NICE (London).

Neff, Gina, and Dawn Nafus. 2016. *Self-Tracking*. Cambridge MA: MIT Press.

Nielsen, Karen Dam. 2015. "Involving Patients with E-health: The Dialogic Dynamics of Information Filtration Work." *Science & Technology Studies* 28 (2): 29–52.

O'Riordan, Kate. 2017. *Unreal Objects: Digital Materialities, Technoscientific Projects and Political Realities*. London: Pluto Press.

Oudshoorn, Nelly. 2012. "How places matter: Telecare technologies and the changing spatial dimensions of healthcare." *Social Studies of Science* 42 (1): 121–142. Accessed February 23, 2021. http://www.jstor.org/stable/23210231.

Oudshoorn, Nelly, and Trevor Pinch. 2003. *How Users Matter: The Co-consctuction of Users and Technology*. Cambridge: MIT Press.

Pantzar, Mika, and Minna Ruckenstein. 2015. "The heart of everyday analytics: emotional, material and practical extensions in self-tracking market." *Consumption Markets & Culture* 18 (1): 92–109. Accessed February 23, 2021. https://doi.org/10.1080/10253866.2014.899213.

Pink, Sarah, and Vaike Fors. 2017. "Self-tracking and mobile media: New digital materialities." *Mobile Media & Communication* 5 (3): 219–238. Accessed February 23, 2021. https://doi.org/10.1177/2050157917695578.

Pink, Sarah, Minna Ruckenstein, Robert Willim, and Melisa Duque. 2018. "Broken data: Conceptualising data in an emerging world." *Big Data & Society* 5 (1). Accessed February 23, 2021. https://doi.org/10.1177/2053951717753228.

Pink, Sarah, Shanti Sumartojo, Deborah Lupton, and Christine Heyes La Bond. 2017. "Mundane data: The routines, contingencies and accomplishments of digital living." *Big Data & Society* 4 (1). Accessed February 23, 2021. https://doi.org/10.1177/2053951717700924.

Piras, Enrico Maria, and Francesco Miele. 2017. "Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship." *Health Sociology Review* 26 (1): 38–53. Accessed February 23, 2021. https://doi.org/10.1080/14461242.2016.1212316.

Rich, Emma, and Andy Miah. 2017. "Mobile, wearable and ingestible health technologies: towards a critical research agenda." *Health Sociology Review* 26 (1): 84–97. Accessed February 23, 2021. https://doi.org/10.1080/14461242.2016.1211486.

Ruckenstein, Minna. 2014. "Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles." *Societies* 4 (1). Accessed February 23, 2021. http://www.mdpi.com/2075-4698/4/1/68.

Ruckenstein, Minna, and Natasha Dow Schüll. 2017. "The Datafication of Health." *Annual Review of Anthropology* 46 (1): 261–278. Accessed February 23, 2021. https://doi.org/10.1146/annurev-anthro-102116-041244.

Schwennesen, Nete. 2019. "Algorithmic assemblages of care: imaginaries, epistemologies and repair work." *Sociology of Health & Illness* 41 (S1): 176–192. Accessed February 23, 2021. https://doi.org/https://doi.org/10.1111/1467-9566.12900.

Sharon, Tamar, and Dorien Zandbergen. 2017. "From data fetishism to quantifying selves: Self-tracking practices and the other values of data." *New Media & Society* 19 (11): 1695–1709. Accessed February 23, 2021. https://doi.org/10.1177/1461444816636090.

Star, Susan Leigh, and Anselm Strauss. 1999. "Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work." *Computer Supported Cooperative Work (CSCW)* 8 (1): 9–30. Accessed February 23, 2021. https://doi.org/10.1023/A:1008651105359.

Urban, Monika. 2017. "This really takes it out of you!: The senses and emotions in digital health practices of the elderly." *Digital Health* 3 (1). Accessed February 23, 2021. https://doi.org/10.1177/2055207617701778.

Weiner, Kate, Flis Henwood, Catherine Will, and Rosalind Williams. 2017. "Self-Monitoring for Health: Questions for an Emerging Field." *SSRN*. Accessed February 23, 2021. http://dx.doi.org/10.2139/ssrn.3051201.

Weiner, Kate, and Catherine Will. 2018. "Thinking with care infrastructures: people, devices and the home in home blood pressure monitoring." *Sociology of Health & Illness* 40 (2): 270–282. Accessed February 23, 2021. https://doi.org/doi:10.1111/1467-9566.12590.

Whitson, JR 2013. "Gaming the quantified self." *Surveillance & Society* 11 (1/2): 163–176.

Will, Catherine M., Flis Henwood, Kate Weiner, and Rosalind Williams. 2020. "Negotiating the practical ethics of 'self-tracking' in intimate relationships: Looking for care in healthy living." *Social Science & Medicine* 266. Accessed February 23, 2021. https://doi.org/https://doi.org/10.1016/j.socscimed.2020.113301.

Williams, Rosalind, Kate Weiner, Flis Henwood, and Catherine Will. 2020a. "Constituting practices, shaping markets: Remaking healthy living through commercial promotion of blood pressure monitors and scales." *Critical Public Health* 30 (1): 28–40. Accessed February 23, 2021. https://doi.org/10.1080/09581596.2018.1497144.

Williams, Rosalind, Catherine Will, Kate Weiner, and Flis Henwood. 2020b. "Navigating standards, encouraging interconnections: Infrastructuring digital health platforms." *Information, Communication & Society* 23 (8): 1170–1186. Accessed February 23, 2021. https://doi.org/10.1080/1369118X.2019.1702709.

Woodward, Sophie. 2015. "The hidden lives of domestic things: accumulations in cupboards, lofts and shelves." In *Intimacies, Critical Consumption and Diverse Economies*, edited by Emma Casey and Yvette Taylor, 216–232. Basingstoke: Palgrave Macmillan.

Woodward, Sophie. 2016. "Object interviews, material imaginings and 'unsettling' methods: Interdisciplinary approaches to understanding materials and material culture." *Qualitative Research* 16 (4): 359–374. Accessed February 23, 2021. https://doi.org/10.1177/1468794115589647.

User-Oriented Innovations: On Cooperative Imagination Spaces in R&D Projects to Support Older Adults in Rural Areas with ICT and Sensor Technology

Claudia Müller and David Struzek

In the development of digital technologies, a large gap persists between the ideas of what is potentially possible with innovative technologies such as complex algorithms, sensor media, adaptive and self-learning systems on the one hand, and the everyday lives of potential users on the other. This holds especially true for older and often non-tech-savvy people, who are considered as an important target group for emerging digital technologies in the health and ageing development environment. The question of how data-intensive technologies that rely on various sensors for collecting data can be designed to find a meaningful fit within the lifeworlds of older people is therefore crucial within applied informatics research and development (R&D) projects.

This is echoed by German and European funding programs that increasingly demand the involvement of user groups in the form of user-centered or Participatory Design for interdisciplinary projects with academia-industry cooperation. In such projects, cooperation is inherently multiple, i.e. project work is distributed across heterogeneous cooperative settings with different stakeholders and/or actors involved in each setting. So far, however, little attention has been paid to how these multiple forms of cooperation between highly diverse stakeholder groups, each with their own different visions, are linked to each other over a project period that usually lasts several years. There seems to be an assumption that it is sufficient to bring together different disciplines such as technology researchers, e.g., sensor technology or pattern recognition experts, with user research experts as well as representatives of the target group. Through the use of user-oriented methods such as being pursued in Participatory Design or in the Living Lab approach, the bridging of the gap between high-tech ideas and everyday worlds seems easy to implement.

In this chapter, we place this gap at the center of consideration and examine "how common goals, means and processes" (Schüttpelz 2017, 24) can be mutually as well as inclusively accomplished across heterogeneous cooperative settings within a project. We present our reflections on the use of Participatory Design and Living Lab methods on the basis of a concrete nationally funded interdisciplinary project from a funding line that aims to develop adaptive, self-learning systems for the support of older people in rural areas.

In this way, we would like to provide reflections on the many facets of framing conditions of cooperation in which data practices are always embedded in R&D projects, but which have hardly been addressed so far. To this end, we follow a practice-based approach of Socio-Informatics, a sub-field of applied computer science that pursues a praxeological foundation of research and design work (Wulf et al. 2018). The socio-informatics view takes a close look at the socio-cultural conditions of emergence and processes in cooperative design projects and thus extends classical concepts of applied computer science, such as methods that are labelled under the umbrella of "user orientation" (Kuutti and Bannon 2014). The praxeological approach resonates with recent investigations into media of cooperation in science, technology and media studies that analyze media not only as means or tools enabling cooperation, but also underline their cooperative production which is seen as an ongoing accomplishment (Schüttpelz 2017, 24). Digital media are not just seen as technological artifacts, but as grounded in practices that span all stages of development and use involving various stakeholder groups. We will argue that the setup of joint spaces for anticipating and imagining future technologies along with interlinked media and data practices is crucial when involving target user groups with little or no previous experiences with digital technologies (Meurer et al. 2018; Gießmann et al. 2019). It is our aim to elaborate on the discourses, methods and different interdisciplinary and intersectoral approaches which all come together in an R&D project and thus have impact on the research designs, the final products as well as the imagination spaces which are being collaboratively produced and sometimes fit together better and sometimes less well.

In the remainder of the chapter, we look at two frequently used "user innovation" methods for the field of IT design to support the home life of older people from a socio-informatics perspective, namely Participatory Design and Living Labs. Then we introduce the case of a concrete R&D project to discuss the challenge of building bridges between high tech visions and development goals for adaptive and self-learning systems and real everyday worlds of older people in rural areas through participatory approaches. Building on this, we point out that the establishment of cooperation spaces and media in such highly complex projects must be broken down to a consistently user- and practice-oriented perspective that must also consider additional elements of embedding, such as offers of appropriation support and engagement for older people that help them to develop an interest in information and communication technologies (ICT) and sensor media and which enable them to participate as competent cooperation partners in the design of media and data practices that are meaningful for them.

R&D Projects in the Field of Information Technology to Support Health and Ageing

Recent demographic changes in Europe such as increasing life expectancy and reduced birth rates are linked to drastic changes in respect of age structures. The number of people aged 80 and over will have doubled by 2025; yet at the same time, the availability of workers in the care sector will be drastically reduced (European Commission 2015a). In its program "Innovation for Active & Healthy Ageing," the European Commission faces these challenges for the future, attributing ICT and sensor technologies a major role in the development of innovative solutions for preventive and curative measures. Novel technologies are seen as a major driver for quality of life and increasing the agency of the elderly in their everyday lives (European Commission 2015b).

ICT and sensor-based systems are seen as having a high potential for securing the quality of life in rural areas (Trapp and Swarat 2015). Rural areas exhibit special aspects of demographic and structural change and are therefore the focus of particular attention in technology projects. Challenges relate in particular to the provision of public services and health care as well as the mobility of older rural residents. This is based on the increasing dismantling of social and institutional infrastructures, such as the decline of church services in villages or the dwindling of social meeting places and facilities for daily local supplies such as shops, restaurants, or pubs.

In the past decades, a lot of funding has been spent on the development of new digital solutions supporting quality of life and care of older adults, but few innovations have been broadly accepted by its targeted end-users thus far (Chung et al. 2016). Research on barriers of technology acceptance is abundant and diverse. A major reason for the lacking uptake of such digital technologies is seen in belated and inadequate user involvement (Mort et al. 2015). As a result, innovations too often do not address end-users' needs and/or challenge daily routines (Fitzpatrick and Ellingsen 2013). They do not match cultural values, psycho-social needs and do not fit into everyday practices and in turn do not become embedded into the social world they were designed to become part of (Procter et al. 2018). The early and consistent integration of end-users is therefore increasingly seen as a mandatory requirement for product innovation and development. European and German funding programs have been adapted accordingly, launching research policies which encourage project designs that follow a more integrated real-world perspective fostering co-creation and participatory research and design, which puts the inclusion of primary, secondary, and tertiary end-user groups at the forefront. However, its implementation is demanding and often hesitantly implemented (Rodriguez et al. 2013; Stubbe 2018). Two prominent methodologies that are widely discussed and implemented are the Living Lab and the Participatory Design approach. In the following, main tenets of these approaches will be introduced, then the challenge of applying those methods in R&D projects with older adults will be highlighted by reporting on a project example aiming at the co-development of digital infrastructures supporting societal participation and inclusion of older citizens in a rural German area. Finally, we will subsume possible solutions that might help in paving a way to push the research on IT support infrastructures for older adults into a more practice- and user-based direction.

User-Oriented Innovation Methodologies in Germany: Participatory Design and Living Labs

Participatory Design of software involving future users originated in Scandinavia (Norway) in the 1970s. Its origin was the cooperation between technology researchers, companies, and trade unions. The common goal was to implement the computerization of workplaces in a democratic way. The Scandinavian approach follows three core principles that still guide Participatory Design today (if taken seriously): 1. democratization, 2. emancipation, and 3. product quality (Bødker and Pekkola 2010). In this sense, a participation-oriented IT development approach that pursues product development for older people should enable cooperation with older people as representatives of the target group at eye level. The design, introduction, and evaluation of the jointly developed devices take place in a joint process between the development team and the project participants or co-researchers (Bratteteig and Wagner 2016).

Living Labs is a methodological approach that clearly puts user-involvement to the fore. The term appeared in the 1990s and has since then been used for a wide range of environments and approaches for ICT innovation and development (Følstad 2009). Living Labs are inter- and transdisciplinary social spaces of co-learning, where people from various backgrounds meet on a regular basis to mutually learn from their respective experiences and co-create new solutions for commonly understood problems (Riva-Mossman et al. 2016). The ultimate goal is to create technological interventions or tools that adapt to and work with already existing (care) structures, (human) resources and practices.

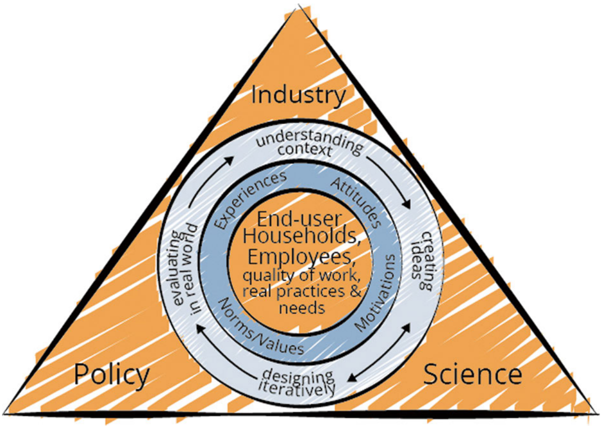
The first Living Labs were artificial laboratories, furnished like regular apartments (Intille et al. 2005; Olivier et al. 2009). Subsequently, researchers and (mainly technology) companies went on to establish Living Labs in real-life environments. This kind of Living Lab can be divided into two sub paradigms: test bed-like settings primarily for evaluation and innovation purposes, and more private, smaller-scaled household settings (Følstad 2009; Ogonowski et al. 2018). These settings seek to provide a frame for cooperation between researchers, companies, governance, and future users in order to create holistic, sustainable and innovative products.

While methods of participant observation have been applied in workplace studies (Suchman 1987; Hughes et al. 1992) since the 1980s, the penetration of private households for technology research purposes emerged only recently when the trends in ubiquitous computing, home automation and smart homes were demanding a deeper understanding of human practices in private environments. This form of Living Lab supported by ethnographic methods such as diary studies, participant observation or cultural probes can provide a huge variety of qualitative data that may support the researchers in understanding certain practices on a very detailed, personal level as well as in identifying general attitudes, problems and needs (Crabtree 1998).

A specific form of Living Lab that fosters real-world and long-term engagement with research participants is the PraxLabs framework (Ogonowski et al. 2018) as a particular form of the Living Lab approach promoted by the EU in its Open Innovation strategy.[1] The PraxLabs framework combines a mix of methods suitable for research in sensitive contexts and with vulnerable research participants, including ethnography (Randall et al. 2007), Participatory Design (Bratteteig and Wagner 2016) and Value Sensitive Design (Friedman et al. 2008). Within the PraxLabs research framework, end-user participation is given in all phases of the development: (a) during pre-studies in order to find and define requirements; (b) in the reflection of early design prototype versions; (c) during the improvement of the usability and meaningfulness of solutions to end-users, (d) when testing in practice to provide continuous feedback during use (Ogonowski et al. 2018) (see figures 1 and 2).

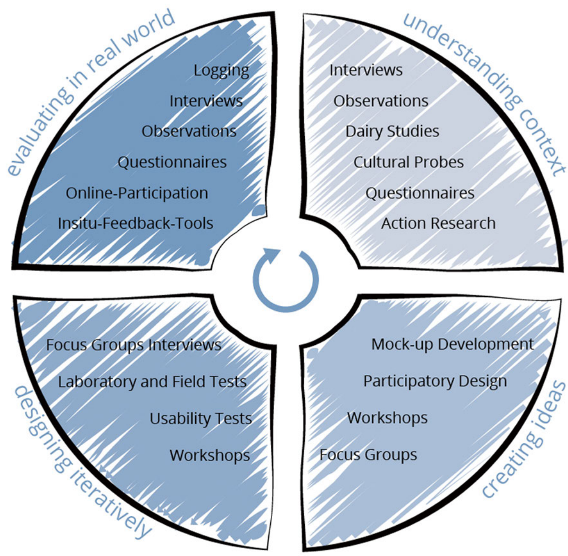
1 https://praxlabs.de/.

Figure 1: PraxLabs approach



Source: Praxlabs.de, Praxlabs Universität Siegen (2021)

Figure 2: Qualitative Methods in Practice-based IT development



Source: Praxlabs.de, Praxlabs Universität Siegen (2021)

The further away an anticipated future technology is from established practices, the bigger the challenges for user- and practice-oriented research projects. As a consequence, particular attention must be paid to the question of how representatives of the intended target groups can be actively involved in the research and development processes, especially if there is little or no prior experience with digital media. This requires a careful and continuous dialogue and mutual learning between project participants and the development team (Meurer et al. 2018). For R&D projects for and with older people, despite the increasing sensitivity to real-world contexts, there are still several hurdles that need more attention in the future (Hornung et al. 2017).

Challenges in Participatory Technology Development For/With Older Adults

Methodologies denoted as "participatory" are currently displayed in highly diverse ways and so are any associated activities being negotiated and implemented. Approaches range from relatively simple lists of methods for user integration in Ambient Assisted Living (AAL) projects (Nedopil et al. 2015) to more in-depth reflections on what participation and co-design imply for collaborative settings between academic actors and participants from application fields, and what conditions need to be met for successful implementations (Vines et al. 2015). Technology research in the field of elderly care must take into account aspects of age and diversity as well as the diversity of contexts and life situations (e.g., complex care arrangements, cohort effects) but also the qualifications and digital experiences of users (e.g., technology biographies) (Meurer et al. 2018). These considerations must be differentiated between preliminary studies, co-design, application testing, evaluation, and implementation as well as the design of sustainable learning and appropriation settings and urgently require further methodological development (Joshi and Bratteteig 2016).

Additionally, the discussion of a good life across all life spans, especially in old age, is of special interest in current research. However, these debates are often primarily framed in terms of a health-economic oriented rationality that privileges technology-driven design ideas which lack the acknowledgement of everyday practices, lifestyles, norms and value sets of older adults. Despite a highly vibrant research landscape in the field of digital systems for older people, studies show that the transfer of research and development results has progressed only moderately so far. An important reason for this is often evident in the way older people are "configured" as future users. Neven and Peine (2017) point out two major deficits in design processes for older people: Firstly, the definition of research objectives is often guided by deficit-oriented images of ageing. Older people are therefore portrayed as passive recipients of technology. As a consequence, secondly, older people appear as mere "research objects instead of as active co-researchers.

In addition, there is a widespread idea of older adults not being capable any more to learn digital practices and of being generally technology-skeptical. These issues point to barriers to access to digital media, ranging from usability issues to stereotypical and images of age and ageing (BMFSFJ 2020). There are numerous findings in the literature that R&D projects are frequently characterized by technology-centricity and stereotyping images of age (Gallistl et al. 2020).

In addition, Bratteteig and Wagner (2012) point out that "user participation has become important not only in the design of IT, but also in areas such as health care, community development and urban planning." Here, one challenge in Participatory Design emerges as central: overcoming the "asymmetry of knowledge" and creating a "symmetry of knowledge" (Fischer 2000) between the designer/developer, who is aware of the "design space," and the users involved, who are aware of the "problem space." What is needed to create this symmetry is a process of mutual learning (see e.g., Hornung et al. 2017) to establish a shared hybrid space or "third space" (Muller and Kuhn 1993) that extends both – the design space and the problem space – towards the design goal.

Project Example: Participatory Design of ICT and Sensor Technologies with Older Adults in a Rural Environment

Cognitive Village (Kurz et al. 2020) was a R&D Project situated at the University of Siegen that aimed at the exploration and development of ICT and sensor technologies for supporting the quality of life and autonomy of older rural residents. There was the basic assumption that new technologies may offer potentials when being embedded in every-day environments and within the local social networks and organizations. Therefore, one of the central project activities aimed at learning from the everyday practices of older adults in the context of their conduct of everyday life in their rural villages for stimulating design ideas for innovative applications. An interdisciplinary research team at the University of Siegen worked together with a group of older residents, with local service providers as well as with representatives of a church community and a technology company building on Living Lab elements and Participatory Design.

There was a huge contrast between the R&D goals in innovative sensor development and the everyday life worlds of the older residents in the village. Thus, setting up a Participatory Design process at eye-level was highly demanding and faced various challenges: How could we spark interest in the local community for some highly abstract sensor technology ideas? And how could we do this from a community-based perspective, addressing different societal groups in the village? For addressing these issues, a set of different strategies has been pursued.

The initial contact with the village residents was made through information meetings with key people from the region. Participants were representatives of the municipalities, local associations and the church community, the resident general practitioner and volunteer operators of a recently opened village shop. These people took on two essential functions. On the one hand, they themselves contributed ideas on how digital technology could possibly contribute to strengthening the quality of life of older residents in the context of village communities. On the other hand, they acted as multipliers and intermediaries for sparking interest and motivation in the older residents to engage with the research project.

In several meetings, possible themes emerged, including the use of apps and a home application for fall prevention. This proved to be an interesting anchor point for the project as a whole that resulted in possibilities for using pattern recognition software later on. Engaging in a project on fall prevention appeared interesting to a group of 15 older (62–85 years) residents who joined follow-up workshops with the research team. We introduced off-the-shelf devices, such as smartphones, tablet PCs and bracelets as well as tracking applications which proved to be a common starting point to be of interest to the local elderly residents, but likewise for the researchers. With regular meetings around movement tracking (e.g., step counters), it was a starting point for engaging with the data collected by these applications and for exploring potential practices based on such data that relate to existing interests and habits.

In order to make good use of the technical system – and also to be able to discuss possible future developments with the researchers later on – it quickly became clear to the participants that digital competence and experience with technologies would be necessary. Thus, there was a high willingness to embark on the path of technology use of smartphones, tablet PCs and fitness bracelets which had been provided by the research team within the framework of so-called appropriation cafés (see figure 3) that took place on a bi-weekly basis between 2 and 2.5 hours.

The appropriation cafés thus became the central place for joint technology exploration, use and reflection, which were flanked by meetings and workshops with other local actors. As the older co-researchers' experience with the tools in everyday life increased, so did their interest and competence to discuss other options for use. On this basis it became only possible to test and discuss versions of the pattern recognition software in the joint meetings.

Figure 3: Appropriation café in action

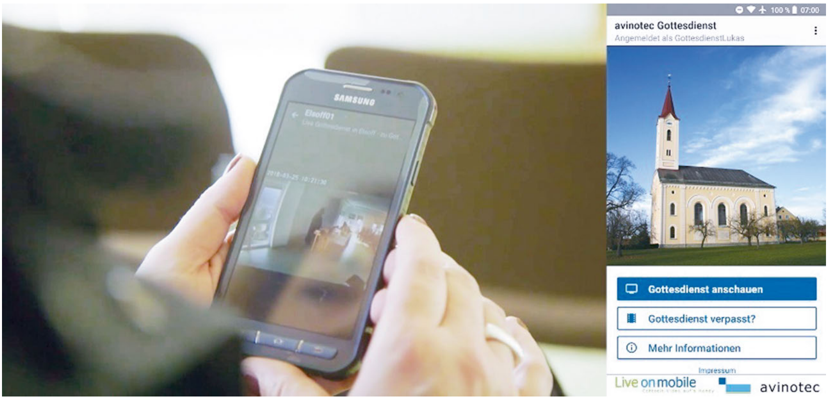
Source: Author's picture

The regular appropriation cafés form a central element of the Participatory Design process established within the Cognitive Village project. Older people want to understand what they can use technology for, so it has to be meaningful for them to embark on the learning journey. However, the creation of meaning does not fall from the sky, but needs certain frameworks, such as the accompanied first steps of using the devices and software and the joint finding of anchor points in the life worlds and interests of the users in conversations and joint explorations with the researchers. Through this dialogue-based joint exploration phase, interests were awakened, motivation strengthened and the necessary media skills for participation in the design process were learned (Struzek et al. 2017).

However, this kind of participatory research is not a linear process when many different groups of actors on the ground as well as an interdisciplinary team work together. A number of ideas were developed, elaborated, discussed and discarded. Some project ideas, however, were implemented in such a way that they are still valid after the end of the project.

After a couple of months, when the constant group of 15 co-researchers had committed themselves as regular research participants, we started to discuss the question of how we could interest more people in the technology exploration process. We were able to create stronger links to the village shop volunteers and the church community. The "core" group started to have fun thinking about how to make the joint technology exploration process more visible to other older village residents. In some following Participatory Design workshops the group developed ideas on how a church camera for streaming Sunday church services and a digital blackboard in the village shop could be implemented and they took part in an iterative prototyping phase. Both digital systems then had been implemented and the group developed ideas and measures for promoting the new tools among other older adults in the village.

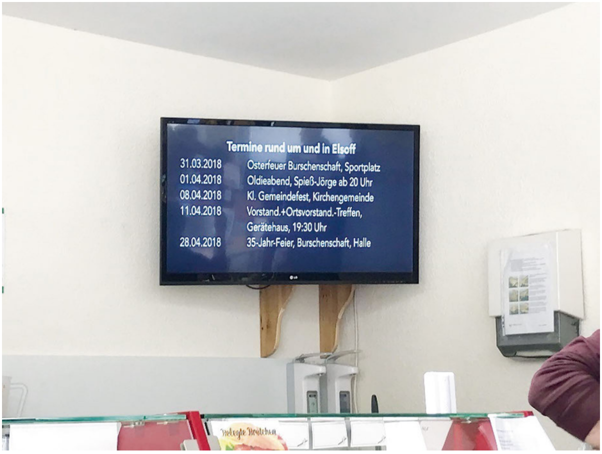
Figure 4: Church camera



Source: „Gut vernetzt im Alter," Arte (2018)

All in all, different lines of development of digital technology had been followed – the "high tech" pattern recognition project as well as the participatory development processes for the more mundane village shop blackboard and the church camera to broadcast Sunday Mass. In the end, this strategy proved successful for setting up a sustainable Participatory Design process which grounded IT development in real-life circumstances of older adults as well as in the social fabric of a local community. The mundane technologies have been handed over to the village shop operators and the parish after the project ended. The part of the pattern recognition project did not end in a market-ready project, which also was not the aim of the R&D funding line. However, the pattern recognition scientists had great opportunities for learning and grounding their algorithm development processes in a real-world setting.

Figure 5: Digital blackboard in the village shop displaying local event dates



Source: Author's picture/aviontec Gottesdiesnt GmbH

Discussion: Grounding Technology Development in Real-World Settings

Participatory Design is time and resource-intensive and thus most often takes place in small-scale settings. Thus, integrating this approach in a Living Lab research design that aims at bringing together a range of different stakeholders in a village poses several challenges.

Firstly, there is the need to bridge between research ideas and goals for "high tech" data-intensive digital technologies (e.g., pattern recognition algorithms) and actual everyday life settings of older adults. How can we talk about future technologies with people who are barely familiar with "simple" smartphones – if at all? How can we at least find some links between algorithm research which most of the time takes place in a lab and real life contexts of older adults which are most of the time not digitized so far? If we wish to take participation and grounding of R&D research in everyday life seriously, we need to be in those contexts and enroll the topic from both sides. The pattern recognition colleagues most of the time stayed in their lab, but from time to time visited our workshops with the older research partners. We as socio-informatics researchers built the bridges between the different stakeholders.

Grounding a technology research project in a local community at first needs to spark interest in the people. Meetings with local "door-openers," i.e. communal actors, engaged people in local associations and the like are important to talk to at first and let them help to get in touch with community residents.

Next, the aspect of "enabling for co-design" (Hornung et al. 2017) is of utmost relevance if the project wishes to invite older residents for a long-term participation. Our approach in organizing meetings with coffee and cake and providing a comfortable atmosphere provides low-threshold opportunities for getting in touch with each other to negotiate a topic that is of interest to the participants from the viewpoint of their individual life-worlds. In addition, bringing in off-the-shelf devices such as smartphones and tablets and slowly trying out applications in relation to their interests helped in finding anchor points in their everyday life that made technology use interesting to them and helped us in turn to gain a better understanding of their interests and needs. The workshops thus served as learning spaces where a dialogic form of learning could take place between "us" and "them." After several months of getting familiar with the devices and software for their own usage interests, amongst others, the movement tracking applications, they felt familiar with the devices and were eager to think about how they can inspire their fellows and friends about the use of these technologies.

The two other systems were simple technology-wise, but rather complex to be implemented in community structures. The church camera and the blackboard served as "boundary objects" to other older residents who had neglected or hesitated to engage with the project and with digital technology in general. In a joint approach with members of the parish and the village shop volunteers, those systems stayed after the project and might spark reflections and conversations about follow-up projects in the village. At least, these off-the-shelf technologies have helped to establish practice-based options of what digital technology could be used for and provide the ground for a next follow-up project.

Conclusion

Participatory Design and user-oriented IT development approaches are very popular today – however they are used in highly differing ways.

Participatory Design work in a Living Lab setting comes with a lot of challenges, which need to be taken seriously, otherwise there would be no chance to truly ground the research in real-life practices and structures.

Introducing off-the-shelf products as a first step for starting a joint exploration and sense-making journey proved successful. In addition, the implementation of technology in community spaces is being seen as a long-term strategy for sparking interest for a further engagement with the digital world by creating opportunities to explore, experience and talk about the tools to other people in the community.

Having said this, it implies to take on a long-term perspective for the development of community-based strategies for technology introduction, beyond individual project time frames (Meurer et al. 2018). Locally based Participatory Design projects should best be understood as just one step in a regional long-term Living Lab collecting a vast range of experiences and feeding them back into both IT development and community development.

It has been shown that the establishment of common data practices with existing technologies is helpful to develop constructive and concrete visions in Participatory Design processes. Those practices should be considered in their diversity and the interconnectedness of disciplinarity, interdisciplinarity and intersectoral imaginative spaces and practice worlds, when aiming at sustainable implementation strategies.

The chapter also showed that an understanding of cooperation as "mutual making of common goals, means and processes" (Schüttpelz 2017, 24) is helpful to build bridges for sustainable technology development for older people in rural areas between high tech visions and current living conditions. In the future, it will only be possible for a high level of technological innovation to reach the everyday worlds of the target groups and to offer added value as successfully implemented socio-technical systems by reconciling the visions and also the cooperation ideas of all stakeholder groups involved in the project. In this context, an integrated perspective on technology and community development on the ground is inevitable.

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

References

BMFSFJ. 2020. "Older People and Digitisation." *BMFSFJ*. Accessed September 9, 2021. https://www.bmfsfj.de/bmfsfj/meta/en/publications-en/older-people-and-digitisation--159710.

Bratteteig, Tone, and Ina Wagner. 2012. "Spaces for Participatory Creativity." *Proceedings of the 11th Biennial Participatory Design*: 105–26. Accessed September 9, 2021. https://doi.org/10.1145/1900441.1900449.

Bratteteig, Tone, and Ina Wagner. 2016. "Unpacking the Notion of Participation in Participatory Design." *Computer Supported Cooperative Work* 25 (6): 425–75. Accessed September 9, 2021. https://doi.org/10.1007/s10606-016-9259-4.

Chung, Chia-Fang, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A. Kientz, and Sean A. Munson. 2016. "Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & So-*

cial Computing, CSCW '16, 770–86. New York, NY: ACM Press. Accessed September 9, 2021. https://doi.org/10.1145/2818048.2819926.

Crabtree, Andy. 1998. "Ethnography in Participatory Design." In *Proceedings of the 1998 Participatory Design Conference: Computer Professionals for Social Responsibility*, 93–105. Stanford, CA.

European Commission. 2015a. "European Summit on Innovation for Active and Healthy Ageing." *European Commission.* Accessed September 9, 2021. https://ec.europa.eu/growth/content/european-summit-innovation-active-and-healthy-ageing-0_en.

European Commission. 2015b. "Growing the Silver Economy in Europe." *European Commission*. Accessed September 9, 2021. https://digital-strategy.ec.europa.eu/en/library/growing-silver-economy-europe.

Fischer, Gerhard. 2000. "Symmetry of Ignorance, Social Creativity, and Meta-Design." *Knowledge-Based Systems* 13 (7): 527–37. Accessed September 9, 2021. https://doi.org/10.1016/S0950-7051(00)00065-4.

Fitzpatrick, Geraldine, and Gunnar Ellingsen. 2012. "A Review of 25 Years of CSCW Research in Healthcare: Contributions, Challenges and Future Agendas." *Computer Supported Cooperative Work* 22 (4): 609–665. Accessed September 9, 2021. https://doi.org/10.1007/s10606-012-9168-0.

Følstad, Asbjørn. 2008. "Living Labs for Innovation and Development of Information and Communication Technology: A Literature Review." *The Electronic Journal for Virtual Organizations* 10 (7): 99–131.

Friedman, Batya, Peter Kahn, Alan Borning, Ping Zhang, and Dennis Galletta. 2006. "Value Sensitive Design and Information Systems." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by Nelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman, 55–95. Accessed September 9, 2021. https://doi.org/10.1007/978-94-007-7844-3_4.

Gallistl, Vera, Rebekka Rohner, Alexander Seifert, and Anna Wanka. 2020. "Configuring the Older Non-User: Between *Research*, Policy and Practice of Digital Exclusion." *Social Inclusion* 8 (2): 233–43. Accessed September 9, 2021. https://doi.org/10.17645/si.v8i2.2607.

Gießmann, Sebastian, Röhl, Tobias, and Ronja Trischler. 2019. *Materialität der Kooperation.* Springer-Verlag. Wiesbaden: Springer VS. Accessed September 9, 2021. https://www.beck-shop.de/giessmann-roehl-trischler-medien-kooperation-media-of-cooperation-materialitaet-kooperation/product/24554447.

Hornung, Dominik, Claudia Müller, Irina Shklovski, Timo Jakobi, and Volker Wulf. 2017. "Navigating Relationships and Boundaries: Concerns around ICT-Uptake for Elderly People." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 7057–69. Denver, Colorado: ACM Press. Accessed September 13, 2021. https://doi.org/10.1145/3025453.3025859.

Hughes, John A., David Randall, and Dan Shapiro. 1992. "Faltering from Ethnography to Design." In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, 115–22. New York, NY: ACM Press. Accessed September 13, 2021. https://doi.org/10.1145/143457.143469.

Intille, Stephen S., Kent Larson, J. S. Beaudin, J. Nawyn, E. Munguia Tapia, and P. Kaushik. 2005. "A Living Laboratory for the Design and Evaluation of Ubiquitous Computing Technologies." *In CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 1941–44. New York, NY: ACM Press. Accessed September 13, 2021. https://doi.org/10.1145/1056808.1057062.

Joshi, Suhas Govind, and Tone Bratteteig. 2016. "Designing for Prolonged Mastery. On Involving Old People in Participatory Design." *Scandinavian Journal of Information Systems* 28 (1): 1–33. Accessed September 9, 2021. https://aisel.aisnet.org/sjis/vol28/iss1/1.

Kurz, Dana, Marcin Grzegorzek, Claudia Müller, and David Struzek. 2020. "Selbstbestimmt im Alter mit neuer Technik voneinander lernen im Forschungsprojekt Cognitive Village – Vernetztes Dorf." *Forschungskolleg Siegen* 1: 15.

Kuutti, Kari, 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*, 3543–52. New York, NY: ACM Press. Accessed September 13, 2021. https://doi.org/10.1145/2556288.2557111.

Meurer, Johanna, Claudia Müller, Carla Simone, Ina Wagner, and Volker Wulf. 2018. "Designing for Sustainability: Key Issues of ICT Projects for Ageing at Home." *Computer Supported Cooperative* Work 27 (3): 495–537. Accessed September 9, 2021. https://doi.org/10.1007/s10606-018-9317-1.

Mort, Maggie, Celia Roberts, Jeannette Pols, Miquel Domenech, and Ingunn Moser. 2015. "Ethical Implications of Home Telecare for Older People: A Framework Derived from a Multisited Participative Study." *Health Expectations* 18 (3): 438–49. Accessed September 9, 2021. https://doi.org/10.1111/hex.12109.

Muller, Michael J., and Sarah Kuhn. 1993. "Participatory Design." *Communications of the ACM* 36 (6): 24–28. Accessed September 9, 2021. https://doi.org/10.1145/153571.255960.

Nedopil, Christoph, Cornelia Schauber, and Glende Sebastian. 2013. "'Guideline the Art and Joy of User Integration in AAL Projects.' White Paper for the Integration of Users in AAL Projects, from Idea Creation to Product Testing and Business Model Development." *AAL Programme.* Accessed September 9, 2021. http://www.aal-europe.eu/wp-content/uploads/2015/02/AALA_Guideline_YOUSE_online.pdf.

Neven, Louis, and Alexander Peine. 2017. "From Triple Win to Triple Sin: How a Problematic Future Discourse Is Shaping the Way People Age with Technology." *Societies* 7 (3): 26. Accessed September 9, 2021. https://doi.org/10.3390/soc7030026.

Ogonowski, Corinna, Timo Jakobi, Claudia Müller, and Jan Hess. 2018. "Praxlabs: A Sustainable Framework for User-Centered Information and Communication Technology Development—Cultivating Research Experiences from Living Labs in the Home." In *Socio-Informatics*, edited by Dave, Randall, Markus Rohde, Kjeld Schmidt, and Volker Wulf, 319–60. Oxford: Oxford University Press. Accessed September 9, 2021. https://doi.org/10.1093/oso/9780198733249.003.0011.

Olivier, Patrick, Guangyou Xu, Andrew Monk, and Jesse Hoey. 2009. "Ambient Kitchen: Designing Situated Services Using a High Fidelity Prototyping Environment." *In Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*, 1–7. New York, NY: ACM Press. Accessed September 13, 2021. https://doi.org/10.1145/1579114.1579161.

Procter, Rob, Joe Wherton, and Trisha Greenhalgh. 2018. "Hidden Work and the Challenges of Scalability and Sustainability in Ambulatory Assisted Living." *ACM Transactions on Computer-Human Interaction* 25 (2): 1–26. Accessed September 9, 2021. https://doi.org/10.1145/3185591.

Randall, Dave, Richard Harper, and Mark Rouncefield, eds. 2007. *Fieldwork for Design: Theory and Practice.* Accessed September 9, 2021. London: Springer. https://doi.org/10.1007/978-1-84628-768-8.

Riva-Mossman, Susie, Thomas Kampel, Christine Cohen, and Henk Verloo. 2016. "The Senior Living Lab: An Example of Nursing Leadership." *Clinical Interventions in Aging* 2016 (11): 255–63. Accessed September 9, 2021. https://doi.org/10.2147/CIA.S97908.

Rodríguez, Hannot, Erik Fisher, and Daan Schuurbiers. 2013. "Integrating Science and Society in European Framework Programmes: Trends in Project-Level Solicitations." *Research Policy* 42 (5): 1126–37. Accessed September 9, 2021. https://doi.org/10.1016/j.respol.2013.02.006.

Schüttpelz, Erhard. 2017. "Infrastructural Media and Public Media." *Media in Action: Interdisciplinary Journal on Cooperative Media*, 1 (1): 13–61. Fundaments of Digitisation. Accessed September 9, 2021. https://doi.org/10.25819/ubsi/7935.

Struzek, David, Martin Dickel, Dave Randall, and Claudia Müller. 2019. "How Live Streaming Church Services Promotes Social Participation in Rural Areas." *Interactions* 27 (1): 64–69. Accessed September 9, 2021. https://doi.org/10.1145/3373263.

Stubbe, Julian. 2018. „Innovationsimpuls Integrierte Forschung." *Diskussionspapier des BMBF-Forschungsprogramms „Technik Zum Menschen bringen."* Accessed September 9, 2021. https://www.technik-zum-menschen-bringen.de/dateien/service/veranstaltungen/diskussionspapier-integrierte-forschung-2018-05-25.pdf.

Suchman, Lucy A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge, MA: Cambridge University Press.

Trapp, Mario, and Gerald Swarat. 2015. "Rural Solutions: Smart Services für ein Land von morgen." *IM+io.* Accessed September 9, 2021. https://www.im-io.de/digitalisierung/rural-solutions-smart-services-fuer-ein-land-von-morgen/.

Vines, John, Gary Pritchard, Peter Wright, Patrick Olivier, and Katie Brittain. 2015. "An Age-Old Problem: Examining the Discourses of Ageing in HCI and Strategies for Future Research." *ACM Transactions on Computer-Human Interaction* 22 (February): 1–27. Accessed September 9, 2021. https://doi.org/10.1145/2696867.

Wulf, Volker, Volkmar Pipek, David Randall, Markus Rohde, Kjeld Schmidt, and Gunnar Stevens, eds. 2018. *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. Oxford, UK: Oxford University Press. Accessed September 9, 2021. https://doi.org/10.1093/oso/9780198733249.003.0011.

Managing Data, Managing Contradictions: Archiving and Sharing Ethnographic Data

Wolfgang Kraus and Igor Eberhard

During the last two decades, the Open Science movement has swept across academia, with massive repercussions for research and publishing. Starting in the natural sciences and engineering, it has brought wide-ranging promises of better science – more transparent, more reliable, more reproducible, more replicable, more efficient, more accountable, more relevant. A basic requirement to make all this possible, the argument goes, is the opening and managing of research data.

While the social sciences and humanities have been slower to embrace the Open Data paradigm – not least because the category of data itself is controversial – the increasing call for Research Data Management (RDM) and open access to research data has become a major concern in all scientific fields (see Allianz 2010), epistemological and methodological differences notwithstanding. The debate on researchers' responsibilities in producing, handling, and sharing their data, under such headings as Open Research Data or FAIR Data (FORCE11 n.d.), has also been taken up in Social and Cultural Anthropology and related disciplines that use qualitative methodologies such as ethnographic research (see, e.g., Imeri 2017; 2019; Mosconi et al. 2019; Pels 2018). In striking contrast to the optimism outlined above, researchers in these fields often perceive the call for Open Data as intrusive, enforced by research policies and funding agencies but difficult to reconcile with research practices based on relations of trust and, often, confidentiality.

Social and Cultural Anthropology is the discipline in which these reservations arguably are most acute, for several reasons. One is a history connected to European expansion and colonialism which has given rise to a heightened awareness of the ethical dimension of the relation between researchers and the people they study. Another is related to its defining methodology, ethnography, whose practitioners consider those being researched not as objects of study but rather as active collaborators in the construction of knowledge. As a major consequence of this, the understanding of research data in ethnography is very different from the one dominating much of the Open Data discourse (see Pels 2018).

Many ethnographers are deeply skeptical about handing over their data to others, and with good reasons (Imeri 2017; 2019, 49 f.). They fear that submitting to the increasing demand for Open Data may have a damaging impact because it risks undermining the social relations between researchers and research subjects that form the basis of ethnographic research. They also fear that it may compromise the ethical standards they aim to uphold in research and publishing. Giving away data means giving up control about what is done with them, while many researchers feel they can never relinquish the responsibility for the ways in which their data are used.

We argue that the skepticism of many ethnographers and qualitative researchers in the social sciences about data management and sharing is best understood as a result of a series of contradictions that have been discussed extensively in recent years but nevertheless remain underappreciated in much of the RDM policy discourse. A clear contradiction exists between Open Data and data protection (with the usual mitigating strategy being pseudonymization). However, the fact that ethnographic data are best when they are context-rich clashes with the ideal of avoiding personal identification, since pseudonymization only works if data are stripped of context information, or not at all. Furthermore, if ethnography is understood as a collaborative production of knowledge, it is contradictory that only one of the two sides involved – the researchers, but not the research subjects – should have control of the data and decide how they are to be managed.

To a large extent these and other contradictions discussed below remain underappreciated because, just like any field of social interaction, academia too is permeated by inequalities of power. One aspect of this fact is that some disciplinary areas have more power to define agendas such as Open Data, and the policies adopted in their interest: natural sciences more than the social sciences and humanities, quantitative methodologies more than qualitative, and so on. To put it bluntly: ranking low in the hierarchy of disciplines, the qualitative social sciences so far have not had much of a say in the Open Science debates. It is therefore of vital importance that their professional associations make themselves heard by formulating and communicating clear positions in the debate on data management and sharing (see, e.g., DGSKA 2019; DGS 2019; EASA n.d.; see also Boog et al. 2018). Furthermore, researchers should discuss and clarify their principles, practices and experiences concerning these issues. The present chapter is meant as a contribution to this rapidly growing body of literature.

The position we take is that, while indeed much of the skepticism is well founded, there is a lot to be gained from preserving and sharing ethnographic research data, if for different reasons and with different priorities than the ones usually invoked in the Open Data discourse. Based on this conviction, we started in 2017 to set up a digital data archive at the University of Vienna, the Ethnographic Data Archive (*eda*), with the aim of developing strategies and competences for the long-term preservation of ethnographic data and for data management in ongoing research.[1] In this chapter we will comment on the issues raised above and discuss our guiding principles and the insights and experiences we gained in the process. While being a small initiative with limited work capacity, we feel that we are doing important, and in the German-speaking world even pioneering, work, but much of it is still provisional and remains in progress (Eberhard and Kraus 2018, 49 f.).

What are Research Data?

A presupposition of the RDM discourse is that all research is data-based. Clearly, this is a matter both of definition and epistemology. In the humanities the relevance of the research data concept is frequently called into question, at least the way it is understood in the natural sciences (e.g., Andorfer 2015; Drucker 2011; Hügi and Schneider 2013). Peter Pels (2018) similarly argues that in anthropology *research materials* only become *data* once they are commodified, and warns that "reductionist definitions of research data may erase the variability of scientific perspectives and research paradigms" (2018, 3).

An additional difficulty is that the term "data" often implicitly refers to digital data only (e.g., Schöch 2013; FU Berlin n.d.). In the most basic logic of research the format of information – say, of an interview recording – is irrelevant. Nevertheless, many of the possibilities and challenges of managing research data being discussed have to do with their digital format and what might be called, their computability. The problem with the narrow understanding of research data as digital is, however, that by definition all digital information consists of data. Therefore, the category of research data runs the risk of becoming meaningless except as a reference to format, and to a field of methodological practices and possibilities (e.g., the Digital Humanities).

This raises the question of how research data in general are defined. Given the diversity of disciplinary practices, RDM policy statements often resort to mere listings of forms of data being used in various fields (e.g., DFG 2015). Such an approach results in a circular argument when treated as a definition (e.g., Allianz n.d.): research data are defined by their role in the research process, while research itself is defined through the systematic use of data.

Real definitions, by making statements about what research data are supposed to be, tend to give more away than these uncommitted formulations. An often-quoted definition originating from the United States Office of Management and Budget reads: "Research data is defined as the *recorded factual material* commonly accepted in the scientific community as necessary to *validate* research findings …"

1 See https://eda.univie.ac.at/.

(OMB 1999, our emphasis). A slightly expanded version has been picked up by several UK institutions (see, e.g., ESPRC n.d., itself quoted by other institutions). Another influential statement defines research data as "Data that are *descriptive* of the research object, or are the object itself" (University of Bath 2011, our emphasis).

We argue that such an understanding of research data is based on an insufficient model of the research process. On the epistemological level, this model assumes that the main characteristic of research data is their ability to document aspects of the real world in a factual or, at least, descriptive sense, independently of the specific research context that has produced them.[2] Although due consideration is sometimes given to disciplinary specificities and differences, the basic model is often that of the natural sciences and rests on data as quantifiable information. If data reflect the real world independently of their research context, then they are unproblematic to reuse in a different context.

Another assumption in the RDM and Open Data discourse combines epistemology with accountability: access to research data may serve "to reproduce and verify the results" of research, as Austria's main funding institution for basic research states in the context of its Open Access Policy (FWF n.d.). The third and related element is the economic rationale: once public money is invested to fund research, the results must be made openly accessible. "Taxpayer-funded research" is an important buzzword here, and one that raises interesting questions concerning the role of national boundaries with regard to access to knowledge (see, e.g., the US Alliance for Taxpayer Access,[3] or the Foreword in UKRI 2016, 2 for a British example). By the same logic, the data collected during research become assets that cannot be owned by researchers. Rather, they must be shared and the public (including other researchers) has a right to access and re-use them.

Taken together, these assumptions imply two sets of ideas that sit rather uncomfortably with the practice and self-understanding of ethnographic research: first, ideals of objectivity and replicability of research, as well as a sharp discontinuity between everyday knowledge/experience and research-based knowledge, both of which we consider epistemologically mistaken; and second, principles of cost efficiency and accountability about whose neoliberal thrust we have serious reservations. Moreover, we argue that ethnographic research concerns a significant category of others, the research subjects, who are absent from this model of research except when being conceptualized in a paternalistic manner as those whose privacy and rights must be protected.

---

2 It is true that the FAIR principles include "a requirement to openly and richly describe the context within which those data were generated," but this mainly serves "to enable evaluation of its [sic] utility" for secondary use (Mons et al. 2017, 52). We use the term context in a wider, social sense and invoke other arguments for its indispensability (see below).

3 https://www.taxpayeraccess.org/.

Characteristics of Ethnography

In order to put the assumptions just outlined into perspective, let us consider the main characteristics and assumptions of ethnographic research. Ethnographic methods are now being employed across a wide range of disciplines and may mean vastly different things (e.g., Kazubowski-Houston and Magnat 2018). We are here primarily referring to the methodological approach that is mainstream in present-day anthropology, but much of our discussion applies to all forms of ethnography and is relevant for other primarily qualitative methods as well.

To avoid being misconstrued, we wish to make it clear that we are not arguing for a separate epistemology for ethnographic research. Rather, we contend that its example serves to highlight the shortcomings of a positivist model of research that underlies much of the RDM and Open Data discourse (Eberhard and Kraus 2018, 48). It does so by providing an extreme case of how research is always embedded in social relations, and therefore always has a context that needs to be taken into account – something that is increasingly being understood in other fields too. As a medical statistician remarks: "The very production of data is … always relevant to its interpretation" (Barrowman 2018).

In anthropology ethnographic research is typically done over extended periods of time in close contact, collaboration, and exchange with research subjects. It is based on communicative relations with those being researched and sees them as active participants rather than passive objects of study. This and the fact that ethnographic research often deals with personal lifeworlds implies important issues of trust and responsibility.

Ethnography is an open-ended methodology relying on a flexible combination of tools. Hence, ethnographers tend to produce varied and multiple forms and formats of data. Whether in analog or in digital formats, ethnographic data tend to be technically diverse. In order to make sense, the various kinds of data produced must be interpreted in relation to each other and to the overall research context and experience. With the establishment of new digital tools, media, and ways of disseminating and sharing ethnographic knowledge, this characteristic has come to be discussed as "multimodal anthropology" (e.g., Collins and Durington 2018) in recent years.

The predominant understanding and practice of ethnographic research as based on collaborative relations between researchers and the people they work with has several fundamental implications for the understanding of research data and for RDM:

(1) Data are not simply "found" or "collected." They are co-constructed in a process of dialog between researchers and research subjects. They do not merely document facts "out there" but are representations that contain the voices and perspectives of both sides involved.

(2) Therefore, ethnographic data cannot simply be owned by researchers (and even less by their institutions). They also belong to the research subjects and their communities, who may have their own interest in the data.

(3) There are no raw, uninterpreted data in ethnography. The dialogic process of making data is by necessity a process of interpretation.

(4) As products of social relations and dialog, ethnographic data are neither objective nor subjective. Both of these notions are predicated on a model of knowledge presuming a clear distinction between the observer and the observed object, a distinction that is neither meaningful nor possible in ethnographic research. Therefore, ethnographic data are never simply "descriptive of the research object" (University of Bath 2011); their primary referent is the ethnographic relation between researchers and research subjects.

(5) Both sides involved are embedded in specific social and cultural circumstances, bringing these into the ethnographic encounter. There is a gradual difference but no discontinuity between ethnographic knowledge and everyday knowledge and experience. Tearing down this positivist division also opens up the path to incorporating multiple and diverse ways of knowing, as in the debate on Indigenous methodologies (e.g., van Meijl 2019).

(6) Ethical considerations take precedence over considerations of efficiency in ethnographic research. The heightened ethical awareness is a consequence of the historical context in which the ethnographic methodology emerged and the responsibility that comes with its practice. However, the "primary ethical obligation shared by anthropologists … to do no harm" (AAA 2012) does not result from a potentially paternalistic and condescending attitude of researchers knowing what may be harmful for research participants. Instead, it is a concomitant of the fact that ethnographers require their research subjects to engage with them in the collaborative production of knowledge.[4]

Our notion that the common distinction between "raw" and "processed" or interpreted data makes no sense for ethnographic data[5] might appear to contradict the position Pels takes when stating: "Anthropologists should … insist on making an epistemological distinction between 'raw' and 'processed data'" (Pels 2018, 4). However, Pels uses these terms in a different and atypical sense. By "processing" he does not refer to the predominant meaning of making data usable for analysis but rather to the task of preparing them for reuse by others – in the sense of stripping

4 It is instructive to compare the ethics declaration of the German Anthropological Association with its sociological counterpart. While the former (Hahn et al. 2008) stresses responsibility and reciprocity, the latter (DGS 2017) gives priority to objectivity.

5 Also see Barrowman (2018) who argues that even in quantitative science there is no such thing as raw data.

them of information that might be critical or sensitive in the hands of secondary users.

We agree that in many cases “[e]xtensive processing of raw materials (beyond mere anonymization) becomes inevitable if others are to reuse them” (Pels 2018, 4). However, we do not agree with Pels’ argument that sharing data as a matter of principle cannot be reconciled with the research subject’s “rightful claims to knowledge shared with researchers” (Pels 2018, 4). It can, provided that research subjects are explicitly conceptualized as forming part of the audience for whose access and reuse data should be prepared and archived. Ideally, they should also be included in the process of selecting data objects for archiving, assigning metadata and meaning to them, and defining access regulations.

Towards an Ethnographic Archive

The idea of setting up an archive for ethnographic data at the University of Vienna first occurred to us when we realized that the Department of Social and Cultural Anthropology was heading towards a major generational transition, with several colleagues bound to retire over the next years. When reflecting on this development, it seemed to make sense to preserve the ethnographic material they had gathered during their research careers, involving them in the process while they were still available. These considerations did not come up in reaction to the increasing call for RDM and open access to research data. Instead, they grew out of our own research experiences and focused on existing data from earlier research projects, mostly in analog formats, that were to be digitally preserved and made accessible for reuse. The insight that it was also necessary to support ongoing research and provide data management expertise followed later from our experience of working with historical ethnographic materials and from our engagement with the Open Research Data debates. Our initiative thus is not representative of the “‘top down’ policy push” that Mosconi et al. identify as a characteristic of Open Science, but rather of what they refer to as the “collegial desire to share data” (Mosconi et al. 2019, 756).

It took several initiatives and a couple of years until the data archive we had in mind was put into practice, first as a collaborative two-year pilot project of the Vienna University Library and the anthropology department. After the successful pilot phase it was made permanent, albeit with still limited staff resources: Igor Eberhard in a part-time position as archive manager and Jasmin Hilbert as a student assistant, with Birgit Kramreither, the head of the Social and Cultural Anthropology Library, functioning as coordinator and Wolfgang Kraus as scientific leader.

While the main assumptions of (mainstream) ethnographic research as outlined above are probably uncontroversial for most anthropologists, our archival activities and the strategies we devised are based on additional assumptions. Our basic premise is that ethnographic data are of intrinsic interest beyond the primary research context, for two main reasons. First, they are complex and rich in ways which are hardly ever fully exploited in the original analysis. This is often a matter not only of complexity but also of sheer quantity (as both present authors have experienced in their own research). Second, since ethnographic data come out of encounters situated in time and space, they are historical by nature. As a consequence of transformation and change over time, they may become interesting and relevant in unforeseeable ways, not only for researchers and their scientific communities, but also for those being researched. Both reasons, we argue, provide good grounds for preserving ethnographic data and for making them accessible and reusable beyond the original research context, and these are entirely unrelated to the rationale of the Open Research Data discourse.

A further – and for qualitative social researchers perhaps obvious – assumption is that data are meaningless without context. From this widely shared conviction, radically different conclusions can be drawn. Hirschauer (2014), for instance, invokes the context-dependence of data such as interview statements to argue against the call for archiving and reusing qualitative social science data. We take the opposite position (even if we agree about the practical challenges involved, and that data archiving must be a matter of responsible choice rather than mandate). Contrary to Hirschauer, we maintain that it is possible to retain enough context in the archiving process to make data relevant for future uses, and have defined strategies to do so.

The notion of context is a key concept in the qualitative social sciences, and particularly so in present-day anthropology. As Dilley states, "stress on context in interpretation is one of [anthropology's] distinguishing features; and it is relied upon as an indispensable part of anthropological method" (2002, 438). Especially after the discipline's shift to midrange theorizing, contextualization has become an important part of anthropological explanation/understanding, based on the "view that context is generated and negotiated in the course of social interaction and exchange" (Dilley 2002, 439). This is not the place to attempt to clarify the general understanding of context in anthropology, except to note that it forms the indispensable background for a more narrow and tangible notion of context, that of research context, which is essential to our approach to archiving.

Ethnographic knowledge is embedded in social relations and in complex corporeal experience. Specific data objects cannot represent this context by themselves, but must be linked back to it in order to make sense and be interpretable. Moreover, ethnographic data tend to be diverse, and different kinds and formats of data must be interpreted in relation to each other and to the overall research context. A guiding principle for our archival activities is therefore that the link between specific data and their research context must be retained and documented as richly as possible. We will outline below how we are trying to achieve this.

The Ethnographic Data Archive (*eda*): Objectives and Strategic Considerations

The *eda* pilot project started in early 2017 and was made permanent in 2019.[6] Beyond creating and maintaining a digital archive of ethnographic data, *eda* aims to develop best-practice models for preserving ethnographic data for reuse. Our work addresses a wide range of challenges such as: (1) Defining archival and metadata strategies and standards adapted to the specificities of ethnographic research; (2) testing and defining best practice digitization workflows; (3) networking and exchange with other data management and archival initiatives in related fields;[7] and (4) identifying the ethical and legal issues involved and proposing solutions. While our current emphasis is on archiving historical ethnographic materials, mostly in analog formats, we aim to develop a comprehensive research data management strategy for anthropology and related fields in the medium to long term future.

One of our guiding principles is that collaboration with researchers makes more sense than the mere administration of legacies. This is confirmed by experiences we made working with materials existing in the departmental ethnographic collection, with often insufficient metadata and context information. More importantly, this conviction is based on our holistic understanding of ethnographic data objects as representing an interactive research process rather than separate aspects of an independent reality. Therefore, all data must be linked to the research settings and the researchers' biographies, thus enabling archive users to take this context into account. Researchers themselves are in a privileged position to accomplish the task of contextualizing and interlinking the data objects in a comprehensive manner. This also includes explication of theoretical and epistemological assumptions and the political context of research, something which is more obvious when dealing with historical data (e.g., material manifesting colonial involvement and/or racist assumptions), but is relevant with all kinds of data.

6 The *eda* website (https://eda.univie.ac.at/) gives an overview of the team, activities and cooperations; the *eda* team can be contacted at: eda.ksa@univie.ac.at.

7 Our national and international partners include CIRDIS/University of Vienna, Institute for Social Anthropology/Austrian Academy of Sciences, Phonogrammarchiv/Austrian Academy of Sciences, Department of Folk Music Research and Ethnomusicology/MDW, Fachinformationsdienst Sozial- und Kulturanthropologie, Qualiservice/University of Bremen.

Another guiding principle is the respect and support for legitimate interests of research subjects and source communities in the data, their protection from harm, and reciprocity. Here again, researchers are better positioned than archivists or future users to assess the interests and risks involved with specific data sets. As Pels notes, it is their "ethical duty to control how research materials 'go public'" (Pels 2018, 5). In a collaborative and dialogic understanding of ethnographic research, research subjects or their descendants and communities should be included in these decisions as far as possible (something we have not yet been able to accomplish).

On the practical level, we aim at sustainability through optimized workflows, the use of appropriate (open) file formats, standardized procedures and metadata, and ongoing quality control. Devising best practice digitization workflows requires a balancing of conflicting demands. The amount of work and cost involved and the required storage space should be kept low, while the technical quality of a digital copy should be such that it can be expected to be taken as an adequate representation of the analogue original even several decades from now. This requirement, together with a high level of technical autonomy (in the sense of not having to hand over materials to external service providers), is the guiding principle in our digitization strategy. Finally, in line with the considerations above, we leave the decision of what to archive to the researchers, while offering our advice (e.g., concerning ethical or legal issues) when being asked for it.

Data Objects

Eda's data objects are archived using PHAIDRA, the "repository for the permanent secure storage of digital assets at the University of Vienna."[8] As noted above, ethnographic data come in multiple formats. In our archiving activities the main focus is on text (e.g., field notes, transcripts and diaries), images (e.g., diapositives, black-and-white negatives and photo prints) and audio recordings (e.g., interviews, narrations, recitations, music). So far, most of the material has been in analog formats, but some has been digital too, sometimes in obsolete file formats that need to be converted to archival formats. Preferred file formats are PDF/A for text, TIFF for images and AIFF or FLAC for audio material; sometimes it makes sense to also archive the original files. We have not yet started working with film and video due

8 See: https://phaidra.univie.ac.at/. Our close cooperation with the PHAIDRA team has proven to be highly productive and a most pleasant experience. Several people have substantially contributed to establishing and developing *eda*, most important among them: Maria Seissl, Susanne Blumesberger, Raman Ganguly, Rastislav Hudak and Claudia Feigl. We are deeply indebted to them for their ongoing support.

to the multiplicity of file formats, the need for data compression, and our lack of technical expertise in this field.[9]

We did extensive tests with various digitization approaches and defined best practices and workflows for various kinds of material. It should be understood that digitization often is not merely a technical procedure but includes a judgement about the relevant aspects of an object. When copying a faded photograph, for instance, we must decide if we are mainly interested in its current appearance – the result of an aging process – or the original information that can be restored by proper illumination and by digital editing, or both. In each case, the optimal digital copy or copies will be different. Once again, the researchers can help to make such decisions.

Our digitizing and archiving workflows are also based on considerations of the relation between analog objects as potential carriers of knowledge and their digital representations, and the transformation from one state to the other. The example of the faded image shows how a conception of what constitutes the data object in relation to the research context must guide the digital representation. Another example is a Compact Cassette containing several field recordings. Is the single isolated recording the object we are interested in, or is it the cassette as an entity representing a specific time span in the field? We have opted for the second as our predominant perspective. As a consequence, we devised a new object category for the PHAIDRA repository, the "container object," which allows us to retain context by retaining the integrity of the analog object.

The container object is a data object consisting of several files that represent the same analog object. In the case of the Compact Cassette just mentioned, the format of the container object makes it possible to retain the original connection between the recordings. In addition, a cassette often comes with more or less consistent metadata in the form of notes written on it or the cardboard insert in the box. In that case, the container object consists of several audio files and photos of the two cassette faces and the insert.[10] Other container objects, for instance, represent photographs from the ethnographic collection of the department with extensive captions on the verso.

9 For those objects that have already been ingested, see https://phaidra.univie.ac.at/search#?owner=ethnograpp95.

10 E.g., https://phaidra.univie.ac.at/o:953026.

Metadata

Descriptive information is rarely neutral. Seemingly innocent labels such as established names attached to collectivities or technical terms could carry problematic perspectives, ethnocentric assumptions, or obsolete theory. For these and related reasons we gave a lot of consideration to metadata-related questions. Our guiding principle in this process was a conception of research data as situated in the particulars of time and space and in the social encounter of fieldwork. The main consequences for our metadata approach are, first, attention to the research and social context of specific data objects, and second, the obligation to include the research subjects and source communities in our potential audience.

For metadata the PHAIDRA repository relies on a linked-data approach combining several established metadata standards, including Dublin Core, BIBFRAME, SKOS and others. We suggested several additional metadata fields that we considered useful, which were then mapped to existing categories. The PHAIDRA team proved extremely flexible and supportive in helping us to develop an *eda*-specific metadata scheme and submit form.[11] The submit form is relatively self-explanatory in order to enable researchers, after a short introduction, to upload their own data and metadata. This option is further facilitated by the possibility to create project-specific metadata templates. However, regardless of how it is done, providing rich metadata for their data objects requires a considerable effort and time investment by researchers.

Given the complexity of many data objects, we introduced a clear distinction within metadata between several object categories, some of which might coincide in the case of a given object. We refer to them as (1) born-digital object, (2) digital copy, (3) first-order analog object, and (4) second-order analog object.

What is meant with these categories can be illustrated with a musical recording. An audio file from a digital recorder is a born-digital object. A tape recording on a Compact Cassette, once digitized, is the digital copy of the audio content of a first-order analog object, the cassette. When the recording is considered a representation of a specific instrument being played, the instrument is the second-order analog object. Each of these categories, as far as they can be distinguished in a given object, requires its own metadata. An analog object in the second sense may also exist in the case of a born-digital object. The *eda* metadata scheme lets us follow this approach, even if a consistent terminology – not necessarily using our working terms, which could be improved – has not yet been implemented.

11 For examples of objects with fairly complete metadata, see https://phaidra.univie.ac.at/o:1069269; https://phaidra.univie.ac.at/o:1048725.

A second innovation is the notion of "context data."[12] While access to objects in PHAIDRA can be restricted or blocked (an indispensable feature for *eda*), a basic given of metadata in PHAIDRA is that they are always public and open, without access restrictions or authorial responsibility and rights. Our interest in making the research context transparent and tangible turned out to be in potential conflict with this policy. Personal information that helps to contextualize data may not be fit for public access. Background information may be too extensive for metadata and require a research effort that needs to be credited, and so on.

What we refer to as context data provides a pragmatic workaround for these issues. Context data are data objects – often in text format (PDF/A) – providing information on other objects that is not suitable for metadata because it is too complex, needs being protected, or requires authorship and copyright because it is based on personal research, interpretation, and opinion. As separate data objects, context data can have access restrictions; they are referenced in the metadata of the objects they help to contextualize. A significant category of context data is information on research projects and ethnographers' biographies and research trajectories, e.g., the biographical interviews that Eberhard conducted with Elke Mader, now retired professor at the Vienna Department of Social and Cultural Anthropology.[13]

Pseudonymizing and Informed Consent

When sharing data, there is a clear contradiction between the interest in retaining as much context as possible in order to keep data meaningful, and the need to protect research subjects' privacy and identity (see, e.g., Cliggett 2016, 243–245). An ethnographer's main ethical responsibility is to protect research subjects from harm and to respect relations of trust and confidentiality established during research. Personal identifiability is in conflict with these obligations in many research contexts.

With regard to personal information only aggregate data can by anonymous in the strict sense. When aggregation is not intended or does not make sense, as in ethnography and other qualitative methodologies, only pseudonymizing is possible, and is routinely practiced. However, in typically small-scale ethnographic research settings it is not enough to suppress or replace names because just about any information can provide identification cues. Trying to remove all of this information means a radical thinning of data and risks making them useless (see Eberhard and Kraus 2018, 45 f.).

12 See Eberhard 2020a, 173 f.

13 See https://phaidra.univie.ac.at/search#?collection=o:1146526.

An established instrument to handle such conflicts is informed consent. This notion itself, however, needs to be adapted to the logic of ethnographic research and to specific research settings. Originating from and modelled on a biomedical research logic, it carries assumptions whose universality must be questioned, such as a deeply individualistic and ethnocentric conception of risk and consent. Even though granting consent is in most cases a personal process, the underlying decision-making is often strongly socially embedded. This can be observed not only in small Indigenous communities but also in nearby and familiar settings such as rural Austria (Kraus and Seiser 2022, 104 f.). Paying attention to these processes makes it clear that in a given context people may value social visibility in a specific community as much as or higher than the more abstract interest in anonymity. How to balance these contradictory values is often a difficult question, and one that can only be answered satisfactorily in dialog with research subjects.

A second assumption is that consent is given beforehand, once and for all, something which may be better suited to protect institutions (universities, funding agencies, repositories, etc.) than those being researched. If research and the social relations enabling it only unfold in the process, how can we give the information in advance to make informed consent possible at all? Hence the notion of "processual consent" (e.g., Rosenblatt 1995, 148 f.) which is however more demanding to implement and difficult to document than the established model, and grants research subjects rights (e.g., revoking consent) that they are often expected to sign away otherwise. If taken seriously, processual consent means that research subjects must also be included in the data management and archiving process as far as possible (see DGSKA 2019, 2).

When archiving existing data, a further problem arises: they often come from research contexts where documenting consent was not yet expected. In such cases researchers may help to disentangle the processual aspects of research and the expectations of research subjects concerning the use of data. Even where consent has been documented, there is a high chance that the online sharing of data was not considered as an option (Zeitlyn 2012, 471). There are no abstract general solutions for these problems; viable compromises must be found for each single case. They may include the necessity of restricting or postponing access to data analogous to archival closure periods. As far as possible, research subjects and/or their communities or descendants should be included in these decisions. Researchers themselves may also have an interest in protecting their privacy, with consequences for access to data.[14]

14 On the issues discussed in this section, see, e.g., Cliggett 2016; Eberhard 2020a; 2020b; Eberhard and Kraus 2018; Imeri 2017; 2018; 2019; Lederman 2016; Zeitlyn 2012).

Perspectives for Further Development

In order to manage these and other challenges in archiving and sharing ethnographic data, solutions, instruments and perspectives still need to be worked out. At this point, we can only briefly mention three of the most pressing problem areas that must be addressed in the future development of *eda*.

One is the need for graded access to data for various categories of users. A considerable part of ethnographic data is not suitable for open access. In order to avoid the necessity of defensively precluding access to all sensitive material – an option that would defeat the short to mid-term purpose of the archive – several graded levels of access to data must be defined and managed. Models of how this can be done have already been developed (e.g., Imeri and Sterzer 2018; Sterzer et al. 2018) and implemented (e.g., at the Qualiservice research data center[15]). In *eda*'s case, this will have to be done at the level of the PHAIDRA repository which does not yet offer the necessary functionality. In addition, it will also require some amount of staff resources to manage access. But this will be an important and necessary step to increase the relevance of the archive.

The second problem concerns the need for standardization of metadata, and particularly the lack of controlled vocabularies, such as thesauri. This is now widely perceived as a pressing issue with major ethical implications, given that established terms, categories and names often stem from colonial situations, contain assumptions of cultural stability that fail to reflect dynamic processes of change and redefinition, or even carry racist and pejorative connotations. Activities to improve this situation are taking place in various contexts, e.g., in the project GND for Cultural Data (GND4C) (DNB 2019). We are in touch with some of these activities, e.g., a project currently under way at the Fachinformationsdienst Sozial- und Kulturanthropologie.[16] We are also planning a compact model project concerning a subcollection of annotated photographs from the departmental ethnographic collection made by Friedrich Dörbeck during the Hydrographic Expedition to Siberia (1902–12).[17] Taking this material as an example, we intend to evaluate metadata, descriptive terms, ethnonyms etc. and compare them with existing thesauri and technical terminology in collaboration with partners in the region.

This project will allow us to demonstrate existing issues in specific detail and suggest solutions. It will also serve as an exercise in trying to involve research subjects or their descendants, regional experts and others close to the field in the processes of describing and contextualizing data objects – an important aim for our future activities, as recommended by Corsín Jiménez (2018, 4) and others.

15 See note 12.

16 https://www.evifa.de/de/ueber-uns/fid-projekte/gemeinsame-normendatei-gnd.

17 https://phaidra.univie.ac.at/detail/o:1165511.

As a logical extension of the ethnographic approach, the use of collaborative methods of knowledge production in order to make sense of objects and information items have increasingly been put into practice in recent years (for a museum-based example, see Scholz 2017a; 2017b). The question of re-contextualizing data and objects through collaboration and of what might be termed Indigenous metadata bring up a third problem area with even more far-reaching challenges than those just mentioned. It is related to the insight in anthropology and beyond that there is no pure, non-situated knowledge. All knowledge is socially and culturally positioned and serves specific interests. Positionalities and interests are diverse and embedded in relations of power and, often, injustice. Based on such a perspective, Indigenous communities have increasingly begun to claim control over knowledge relating to themselves. Much of this debate revolves around the notion of Indigenous Data Sovereignty (IDS) (Kukutai and Taylor 2016; Walter et al. 2021), i.e. "the right of Indigenous peoples to govern the collection, ownership, and application of data about Indigenous communities, peoples, lands, and resources" (Rainie et al. 2019, 301).

The IDS advocacy and activism movement recently proposed the "CARE Principles for Indigenous Data Governance" to complement the FAIR Principles (RDA IG 2019; Carroll et al. 2020). They aim to exploit the momentum of the Open Data movement to further IDS while simultaneously addressing some of its issues as seen from Indigenous perspectives. CARE stands for "Collective Benefit, Authority to Control, Responsibility, and Ethics" (RDA IG 2019). While Open Data agendas, including the FAIR Principles, typically argue in terms of abstract advantages for science and society in the singular, the CARE Principles stress interests of and benefits for specific communities, acknowledging the "power differentials and historical contexts" that are ignored in the Open Data discourse (RDA IG 2019).

With its focus on communities, collective benefit and diversity, the IDS movement provides a significant and important corrective to the individualist assumptions dominating much of the debate on Open Data and Research Data Management, e.g., its understanding of ethics. At the same time, however, the simplistic dichotomy of "Indigenous" and "mainstream" principles or values employed by some of its main proponents (e.g., Carroll et al. 2020) risks reproducing modernist dualisms, glossing over the diversity within and between Indigenous communities which may manifest itself in equally diverse interests.

Nevertheless, these and other forms of claims to control over knowledge by those being researched[18] are important and productive from an ethnographic perspective. The questions they raise are not only challenging but also enriching for

18 E.g., the "Traditional Knowledge Labels" aiming to define "attribution, access, and use rights for Indigenous cultural heritage" according to community-specific understanding and values (https://localcontexts.org/), which cannot be discussed here for reasons of space.

initiatives such as *eda* when it comes to elaborating and defining the categories of data ownership, control, access, and licensing.

Conclusion

The main assumption underlying *eda*'s activities is that ethnographic data constitute historically situated representations of aspects of a world in flux. As such they do have a value beyond the primary research context and should be preserved and shared. However, the dialogic character of ethnography and the access ethnographers gain to personal life-worlds raise important issues around confidentiality, privacy, and reciprocity.

The management and archiving of ethnographic research data is a highly dynamic field with many challenges and contradictions. Such contradictions cannot be resolved on the level of principles. Nevertheless, it is possible to find pragmatic but ethically sound compromises that enable us to archive and share (with various forms of restrictions) ethnographic data without harming our research subjects. After all, the situation is not unlike the challenges that arise in the process of ethnographic writing and publishing – challenges for which we routinely find pragmatic single-case solutions.

However, these inherent tensions require additional measures that *eda* will have to deal with in the future, and before archiving more sensitive data. In order to strengthen the collaborative aspect of data management, we need models of, and experiences with, the integration of research subjects into all steps of the archiving process. Finally, in order to attain our goal of developing a comprehensive research data management strategy for anthropology and related fields, we must engage with current data practices in research. All of this is difficult to accomplish with the current limited staff. Nevertheless, as a small initiative and in a field where the questions still clearly outweigh the solutions, we feel that we have already achieved a lot.

Peter Pels suggests there is more than one "reason to consider social science data as indigenous or global heritage" (Pels 2018, 3). In this perspective the preservation and sharing of ethnographic data is an ethical obligation. In a similar stance, the American Anthropological Association lists as one of the main ethical principles for anthropologists to "protect and preserve [their] records" (AAA 2012). In a collaborative and dialogic understanding of ethnography, this logically entails the inclusion of research subjects and their communities in the process in order to safeguard data that comply with both FAIR and CARE Principles for the benefit of all interested parties.

References

AAA. 2012. "Principles of Professional Responsibility." *American Anthropological Association*. Accessed October 30, 2021. http://ethics.americananthro.org/category/statement/.

Allianz der deutschen Wissenschaftsorganisationen. 2010. "Grundsätze zum Umgang mit Forschungsdaten." *Gfz-Potsdam.de*. Accessed October 30, 2021. https://doi.org/10.2312/ALLIANZOA.019.

Allianz der deutschen Wissenschaftsorganisationen (n.d.). "Forschungsdaten." *allianzinitiative.de*. Accessed October 30, 2021. https://www.allianzinitiative.de/archiv/forschungsdaten/.

Andorfer, Peter. 2015. "Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung." *DARIAH-DE Working Papers 14, Göttingen 2015*. Accessed October 30, 2021. http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-7-2.

Barrowman, Nick. 2018. "Why Data Is Never Raw." *The New Atlantis* 56. Accessed October 30, 2021. https://www.thenewatlantis.com/publications/why-data-is-never-raw.

Boog, Igor, Henrike Florusbosch, Zane Kripe, Tessa Minter, Peter Pels and Metje Postma. 2018. "Data Management for Anthropologists and Ethnographers: A Position Paper." *Social Anthropology/Anthropologie Sociale* 26 (3): 7–10. Accessed October 30, 2021. https://doi.org/10.1111/1469-8676.12526.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, Kay Raseroka, Desi Rodriguez-Lonebear, Robyn Rowe, Rodrigo Sara, Jennifer D. Walker, Jane Anderson and Maui Hudson. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19 (1): 43. Accessed October 30, 2021. https://doi.org/10.5334/dsj-2020-043.

Cliggett, Lisa. 2016. "Preservation, Sharing, and Technological Challenges of Longitudinal Research in the Digital Age." In *eFieldnotes: The Makings of Anthropology in the Digital World*, edited by Roger Sanjek and Susan W. Tratner, 231–50. Philadelphia: University of Pennsylvania Press. Accessed October 30, 2021. https://doi.org/10.9783/9780812292213-014.

Collins, Samuel G. and Matthew S. Durington. 2018. "Multimodality." In *The International Encyclopedia of Anthropology*, edited by Hilary Callan. Wiley Online Library. Accessed October 30, 2021. https://doi.org/10.1002/9781118924396.wbiea2304.

Corsín Jiménez, Alberto. 2018. *A Data Governance Framework for Ethnography v. 1.0*. Spanish National Research Council, November 2018. Accessed October 30, 2021. http://hdl.handle.net/10261/172227.

DFG. 2015. "Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten." *Deutsche Fördergemeinschaft.* Accessed October 30, 2021. http s://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.

DGS. 2017. "Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS)." *Deutsche Gesellschaft für Soziologie.* Accessed October 30, 2021. https://soziologie.de/dgs/ethik/ethik-kodex.

DGS. 2019. "Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie. Stellungnahme des Vorstands und Konzils der DGS." *Deutsche Gesellschaft für Soziologie.* Accessed October 30, 2021. https://soziologie.de/aktuell/news/bereitstellung-und-nachnutzung-von-forschungsdaten-in-der-soziologie.

DGSKA. 2019. "Position Paper on the Handling of Anthropological Research Data, German Anthropological Association (GAA)." *Deutsche Gesellschaft für Sozial- und Kulturanthropologie.* Accessed October 30, 2021. https://www.dgska.de/wp-content/uploads/2019/12/Positionspapier_Bearbeitet-fu%CC%88r-MV_EN_29.11.2019.pdf.

DNB. 2019. "GND für Kulturdaten (GND4C)." *Deutsche Nationalbibliothek.* Accessed October 30, 2021. https://www.dnb.de/DE/Professionell/ProjekteKooperationen/Projekte/GND4C/gnd4c.html.

Dilley, R. M. 2002. "The Problem of Context in Social and Cultural Anthropology." *Language & Communication* 22 (4): 437–456. Accessed October 30, 2021. https://doi.org/10.1016/S0271-5309(02)00019-8.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5 (1). Accessed October 30, 2021. http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

EASA. "EASA's Statement on Data Governance in Ethnographic Projects." *European Association of Social Anthropologists*. Accessed October 30, 2021. https://www.easaonline.org/downloads/support/EASA%20statement%20on%20data%20governance.pdf.

Eberhard, Igor. 2020a. "Der Kontext bestimmt alles: Kontextdaten und Containerobjekte als Lösungsmöglichkeiten für den Umgang mit sozialwissenschaftlichen qualitativen Daten. Erfahrungen aus dem Pilotprojekt 'Ethnographische Datenarchivierung' an der Universitätsbibliothek Wien." *ABI Technik* 40 (2): 169–76. Accessed October 30, 2021. https://doi.org/10.1515/abitech-2020-2007.

Eberhard, Igor. 2020b. "Herausforderung ethnographische Daten: Erfahrungen und Ergebnisse aus dem Pilotprojekt Ethnographische Datenarchivierung an der Universität Wien." In *Künstliche Intelligenz in Bibliotheken: 34. Österreichischer Bibliothekartag Graz 2019*, edited by Christina Köstner-Pemsel, Elisabeth Stadler and Markus Stumpf, 261–76. Graz: Unipress Verlag. Accessed October 30, 2021. https://doi.org/10.25364/guv.2020.voebs15.20.

Eberhard, Igor and Wolfgang Kraus. 2018. “Der Elefant im Raum: Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien.” *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (1): 41–52. Accessed October 30, 2021. https://doi.org/10.31263/voebm.v71i1.2018.

EPSRC. “EPSRC Policy Framework on Research Data: Scope and Benefits.” *Engineering and Physical Sciences Research Council*. Accessed October 30, 2021. https://epsrc.ukri.org/about/standards/researchdata/scope/.

FORCE11. “Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing Version B1.0.” Accessed October 30, 2021. *force11*. https://www.force11.org/fairprinciples.

FU Berlin. “Was sind Forschungsdaten?” *Freie Universität Berlin.* Accessed October 30, 2021. https://www.fu-berlin.de/sites/forschungsdatenmanagement/ueber-forschungsdaten/forschungsdaten/index.html.

FWF. “Open Access to Research Data.” *Der Wissenschaftsfonds*. Accessed October 30, 2021. https://www.fwf.ac.at/en/research-funding/open-access-policy/open-access-to-research-data/.

Hahn, Hans Peter, Annette Hornbacher and Michael Schönhuth. 2009. “‘Frankfurter Erklärung’ zur Ethik in der Ethnologie.” *Deutsche Gesellschaft für Völkerkunde*. Accessed October 30, 2021. https://www.dgska.de/wp-content/uploads/2016/07/DGV-Ethikerklaerung.pdf.

Hirschauer, Stefan. 2014. “Sinn im Archiv? Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung.” *Soziologie* 43 (3): 300–312. Accessed October 30, 2021. https://publikationen.soziologie.de/index.php/soziologie/article/view/795.

Hügi, Jasmin and René Schneider. 2013. *Digitale Forschungsinfrastrukturen in den Geistes-und Geschichtswissenschaften.* Genève: Haute école de gestion de Genève. Accessed October 30, 2021. https://doc.rero.ch/record/31535.

Imeri, Sabine. 2017. “Open Data? Zum Umgang mit Forschungsdaten in den ethnologischen Fächern.” In *E-Science-Tage 2017: Forschungsdaten managen*, edited by Jonas Kratzke and Vincent Heuveline, 167–78. Heidelberg: heiBOOKS. Accessed October 30, 2021. http://doi.org/10.11588/heibooks.285.377.

Imeri, Sabine. 2018. “Archivierung und Verantwortung: Zum Stand der Debatte über den Umgang mit Forschungsdaten in den ethnologischen Fächern.” In *Working Paper Series des Rates für Sozial- und Wirtschaftsdaten: Archivierung und Zugang zu Qualitativen Daten* edited by Regina T. Riphahn, Gert G. Wagner und Heike Solga, 69–79. Working Paper 267. Berlin: Rat für Sozial- und Wirtschaftsdaten. Accessed October 30, 2021. https://doi.org/10.17620/02671.35.

Imeri, Sabine. 2019. “‘Open Data’ in den ethnologischen Fächern: Möglichkeiten und Grenzen eines Konzepts.” In *Forschungsdesign 4.0 – Datengenerierung und Wissenstransfer in interdisziplinärer Perspektive*, edited by Jens Klingner and Merve

Lühr, 45–59. Dresden: Institut für Sächsische Geschichte und Volkskunde. Accessed October 30, 2021. https://doi.org/10.25366/2019.07.

Imeri, Sabine and Wjatscheslaw Sterzer. 2018. "Schwierige Daten: Forschungsdatenmanagement qualitativer Daten am Beispiel der ethnologischen Fächer" *Präsentation für 107. Deutschen Bibliothekartag in Berlin 2018*. Accessed October 30, 2021. https://nbn-resolving.org/urn:nbn:de:0290-opus4-36849.

Kazubowski-Houston, Magdalena and Virginie Magnat, eds. 2018. *Special Issue: The Transdisciplinary Travels of Ethnography*. *Cultural Studies ↔ Critical Methodologies* 18 (6). Accessed October 30, 2021. https://journals.sagepub.com/toc/csca/18/6.

Kraus, Wolfgang and Gertraud Seiser. (2022). "Lokalgeschichte und Identität: Eine sozialanthropologische Lehrforschung als partizipatives Projekt mit Museum und Gemeinde." In *Geschichte vor Ort: Salzburger Museen und Sammlungen*, edited by Dagmar Bittricher, Andrea Dillinger and Martin Knoll. Salzburg: Edition Tandem.

Kukutai, Tahu and John Taylor, eds. 2016. *Indigenous Data Sovereignity: Toward an Agenda*. Canberra: Australian National University Press. Accessed October 30, 2021. http://doi.org/10.22459/CAEPR38.11.2016.

Lederman, Rena. 2016. "Archiving Fieldnotes? Placing 'Anthropological Records' Among Plural Digital Worlds." In *eFieldnotes: The Makings of Anthropology in the Digital World*, edited by Roger Sanjek and Susan W. Tratner, 251–71. Philadelphia: University of Pennsylvania Press. Accessed October 30, 2021. https://doi.org/10.9783/9780812292213-015.

Meijl, Toon van. 2019. "Doing Indigenous Epistemology: Internal Debates About Inside Knowledge in Māori Society." *Current Anthropology* 60 (2): 155–73. Accessed October 30, 2021. https://doi.org/10.1086/702538.

Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos and Mark D. Wilkinson. 2017. "Cloudy, Increasingly FAIR: Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud." *Information Services & Use* 37 (1): 49–56. Accessed October 30, 2021. https://doi.org/10.3233/ISU-170824.

Mosconi, Gaia, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn and Volkmar Pipek. 2019. "Three Gaps in Opening Science." *Computer Supported Cooperative Work* 28 (3–4): 749–789. Accessed October 30, 2021. https://doi.org/10.1007/s10606-019-09354-z.

OMB. 1999. "Office of Management and Budget Circular A-110." *Whitehouse.gov*. Accessed April 04, 2022.
https://www.whitehouse.gov/wp-content/uploads/2017/11/Circular-110.pdf.

Pels, Peter. 2018. "Data Management in Anthropology: The Next Phase in Ethics Governance?" *Social Anthropology/Anthropologie Sociale* 26 (3): 1–6. Accessed October 30, 2021. Accessed October 30, 2021. https://doi.org/10.1111/1469-8676.125 26.

Scholz, Andrea. 2017a. "Sharing Knowledge as a Step toward an Epistemological Pluralization of the Museum." *Museum Worlds: Advances in Research* 5 (1): 133–148. Accessed October 30, 2021. https://doi.org/10.3167/armw.2017.050112.

Scholz, Andrea. 2017b. "'Wissen teilen' als postkoloniale Museumspraxis – Ein Kooperationsprojekt zwischen der Universidad Nacional Experimental Indígena del Tauca (Venezuela) und dem Ethnologischen Museum Berlin." *Sociologus* 67 (1): 59–82. Accessed October 30, 2021. https://doi.org/10.3790/soc.67.1.59.

Schöch, Christof. 2013. "Big? Smart? Clean? Messy? Data in the Humanities." *Journal of Digital Humanities* 2 (3). Accessed October 30, 2021. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/#big-smart-clean-messy-data-in-the-humanities-n-6.

Sterzer, Wjatscheslaw, Sabine Imeri and Matthias Harbeck. 2018. "Zugriff auf ethnologische Forschungsdaten: Anforderungen und Lösungen" *Poster für 107. Deutscher Bibliothekartag in Berlin 2018*. Accessed October 30, 2021. https://nbn-resolving.org/urn:nbn:de:0290-opus4-157193.

Rainie, Stephanie Carroll, Tahu Kukutai, Maggie Walter, Oscar Luis Figueroa-Rodriguez, Jennifer Walker and Per Axelsson. 2019. "Indigenous Data Sovereignty." In *The State of Open Data: Histories and Horizons*, edited by Tim Davies, Stephen B. Walker, Mor Rubinstein and Fernando Perini, 300–319. Cape Town: African Minds and Ottawa: International Development Research Centre.

RDA IG. 2019. Research Data Alliance International Indigenous Data Sovereignty Interest Group. "CARE Principles for Indigenous Data Governance." *The Global Indigenous Data Alliance*. Accessed October 30, 2021. https://www.gida-global.org/care.

Rosenblatt, Paul C. 1995. "Ethics of Qualitative Interviewing with Grieving Families." *Death Studies* 19 (2): 139–55.

UKRI. 2016. "Concordat on Open Research Data." *UK Research and Innovation*. Accessed October 30, 2021. https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf.

University of Bath. 2011. "ERIM Terminology V4." *Wiki.bath.ac.uk*. Accessed October 30, 2021. https://wiki.bath.ac.uk/display/ERIMterminology/ERIM%20Terminology%20V4.

Walter, Maggie, Tahu Kukutai, Stephanie Russo Carroll and Desi Rodriguez-Lonebear, eds. 2021. *Indigenous Data Sovereignity and Policy*. Abingdon, Oxon: Routledge. Accessed October 30, 2021. https://library.oapen.org/handle/20.500.12657/42782.

Zeitlyn, David. 2012. "Anthropology in and of the Archives: Possible Futures and Contingent Pasts. Archives as Anthropological Surrogates." *Annual Review of Anthropology* 41: 461–80. Accessed October 30, 2021. https://doi.org/10.1146/annurev-anthro-092611-145721.

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

Gaia Mosconi, Helena Karasti, Dave Randall and Volkmar Pipek

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

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

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

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

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

Data as Matter of Care

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

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

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

Challenges for Qualitative Data Sharing

Data sharing and consequently data reuse have been extensively addressed (Heaton 2008; van den Berg 2008; Faniel and Jacobsen 2010). The vast part of the literature, however, deals with practices embedded in the natural and applied sciences. Our matter of care, however, is the additional complexity entailed in the management of qualitative data, where most of the challenges can be characterized as epistemological, methodological, and ethical in nature. For qualitative data, paying attention to the context of their collection and possible re-use becomes an overarching concern. However, what context is, and how to describe it, is non-trivial (Moore 2006). Context determines whether something can be viewed as data or metadata and the "degree to which those contexts and meanings can be represented influences its transferability" (Borgman et al. 2018). Data loses meaning when removed from the original contexts, packaged in repositories, and disengaged from the knowledge and expertise of the researchers who performed the study (Walters 2009). When dealing with qualitative data we need to recognize the essentially reflexive character of data and that it is often rich with personal content (Tsai et al. 2016). Ethnographic approaches are generally based on a relationship of trust between researchers and participants, often in sensitive domains. This leads us to a consideration of the ethical challenges, where protecting the privacy of participants typically is one of the central aims (for more details see contribution by Kraus and Eberhard in this volume, and Eberhard and Kraus 2018).

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

Data Storytelling: Guiding Principles

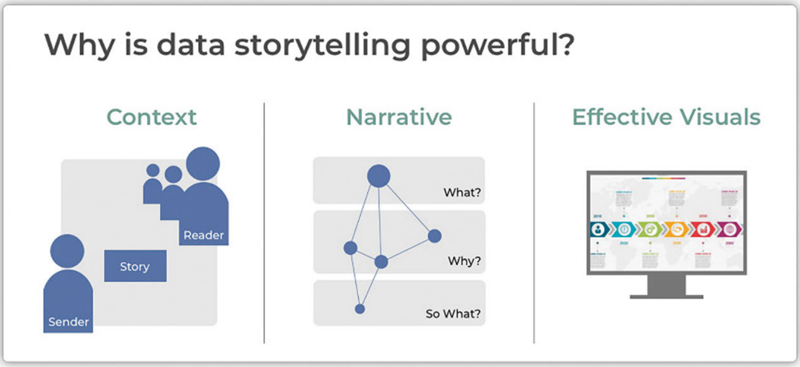
The social sciences and humanities have long stressed the role that narrative plays in human life, in education and in research. As Game and Metcalfe argue:

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

Social scientists tell stories for a range of different purposes. In doing so, they attempt to contextualize the "data" that they work with. They do so largely for analytic purposes. In relation to this, and to return to the question of what context is and how to describe it, there is a difference between context as an analytic construct – something that researchers, curators, etc. define – and something that emerges in and is enacted by the work of the participants. Put simply, context in this view has no existence outside of the way in which it is ongoingly constructed by participants to an activity. Data, in other words, is a process of enactment. Digital storytelling, we want to argue, is a useful means to reconstruct what has previously been constructed or enacted.

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

Figure 1: Main principles of Data Storytelling.



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

As shown in the picture, three main principles summarize what data storytelling is about and how to achieve it: 1) explaining the context; 2) identifying a coherent narrative; 3) working on effective visualization. In data storytelling, the second principle, *narration,* is a crucial element. A narrative can, additionally, have emotional elements. A story has a beginning and an end, it has a goal, sometimes a moral, and, obviously, a story has an audience. Narrative helps to "share norms and values, develop trust and commitment, share tacit knowledge, facilitate unlearning, and generate emotional connections" (Soule and Wilson 2002). The third principle is related to *effective visuals*. As Lee et al. (2015) suggest, relatively little attention has been paid in the visualization literature to the ways in which the stories in question are actually crafted.

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

Background and Approach

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

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

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

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

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

Grounding the Concept in Practices

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

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

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

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

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

The above vignettes point to the way in which a *storytelling approach* to data curation can be called into action, one which is more aligned with researchers' practices, and as possible inspiration to organize the heterogeneous data and to support collaborative data sense-making. In the following, we demonstrate how the Data Story is envisaged to work by showing the design sketches of the low-fidelity prototype we have developed so far. We will then discuss more extensively the idea of selective care that it affords.

The Data Story Process and its Components

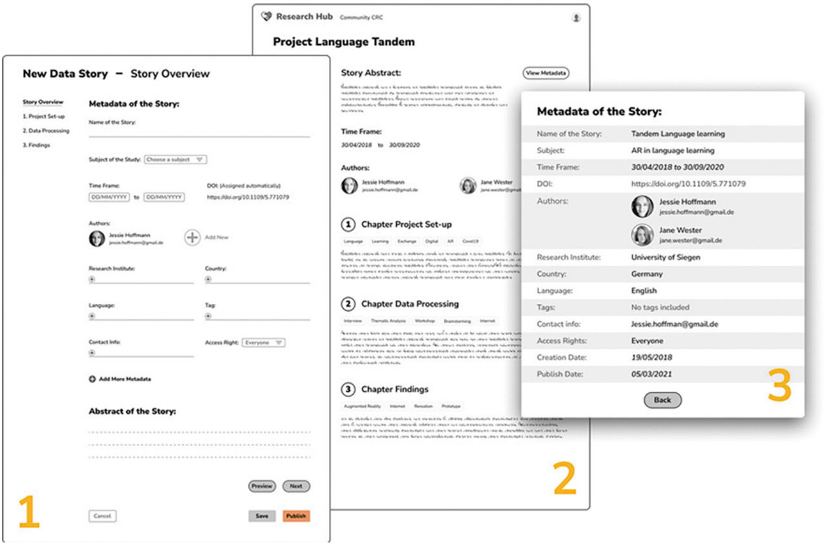
This design concept is meant to be an organizing device in support of (collaborative) storytelling practices as a major component of data analysis and sense making. By engaging with its process and its interactive interface researchers will have the opportunity to perform data curation practices resulting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

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

(0) Overview Screen

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

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



(1) The Project Set-Up Chapter

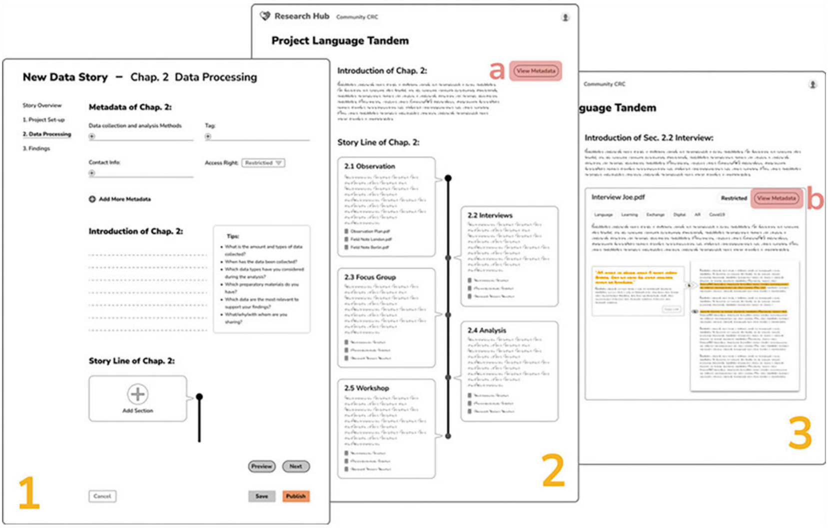
The project set-up chapter introduces the overall story outline, in order to provide an understandable context for the study. Information related to the research field, topic, and research questions of the study, as well as methods used, a short summary about the motivation and aim of the study can be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

(2) The Data Processing Chapter

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

The chapter provides the possibility to create sub-sections which categorize and group data, based on the data type, to ease navigation through it. It is advised to create and fill the sub-sections with relevant data in a way that supports the storyline and sequence of the story. Moreover, this chapter creates a storyline by ordering the created sub-sections sequentially. Authors of the stories will have the ability to relocate the created sub-section if necessary by dragging it to the desired location on the storyline.

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



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

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

(3) The Findings Chapter

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

Supporting Processual Workflows: Plugin Solution

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

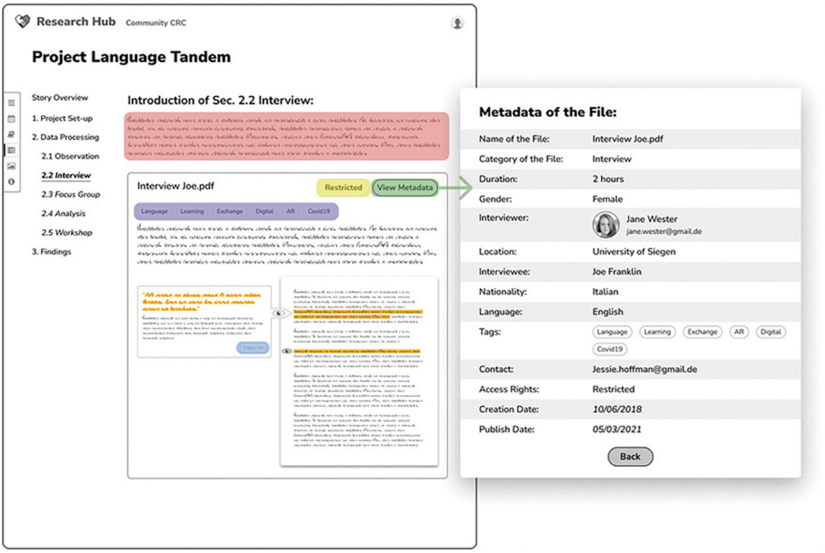
Publishing: DOI and Accessibility Rights

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

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

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

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



In our view, the Data Story should be promoted as a new publication format that is centered around relevant data points. Data Stories can act as an intermediate format between a larger dataset to be stored and secured in long-term archives and the official publications (paper, books etc.). Data Story could offer insights into the content of a dataset but also offering some reflections on the data that might not be part of the final publications. By promoting a Data Story as a new publication format that can be cited, researchers will have the incentives to actually engage with this type of work and get rewarded for this additional effort. For being a Data Story an additional step is important so that researchers will get compensated for this work. By envisioning an accessible open link, Data Stories can circulate freely through the web, and can act as entry points for engagement with the data that have been collected.

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

Data Story as an Act of Selective Care

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

We take on board the injunction of van Es and Schäfer that, "[r]ather than import questions and methods from the hard sciences, we must develop our own approaches and sensitivities in working with data that will reflect the humanities' traditions" (2017, 16). The authors here include a call for action, inviting humanities scholars to develop their own research questions and methods to stay consistent with their epistemological positions. We have shown how we might translate these ideas to the field of Research Data Management and curation. If solutions to data sharing and curation need to be found, as expected and demanded by funding agencies, then we argue, those technical solutions, tools or infrastructures will need to embrace and embed in the design cultural values, methodological practices and epistemological understandings of the communities they are designed for. In doing so, we again connect to the concept of care as pushed forward by Bellacasa (2011): "… representing matters of fact and sociotechnical assemblages as matters of care is to intervene in the articulation of ethically and politically demanding issues. The point is not only to expose or reveal invisible labors of care, but also to generate care" (Bellacasa 2011, 94). We discuss below two lines of argument in which we explicit how the Data Story reveals the invisible labor of data care while at the same time generating care for both the data producer and the data re-user.

Complementing Metadata Standards with a Story

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

The Data Story interface makes visible the act of care by articulating the tasks of data care needed in order to organize the data, retrieve them, present them, share them, and possibly reuse them. In fact, it provides every chapter with the option to annotate, tag and add metadata. The Data Story suggests metadata (i.e.: the Dublin Core or DDI) as the standards source for elements set. They can, however, be adapted quickly and added as new folksonomy. In this way, metadata are treated as "living things" that can grow and develop based on a bottom-up understanding. As mentioned earlier, the Data Story invests noticeable effort in bringing the data and its metadata together by integrating many of the important metadata fields in its interface in a way that makes metadata an important pillar of the story narrative and driver of discussions. It promotes data literacy and awareness, as it is an opportunity for researchers to learn about the role of metadata but also put it into question and adapt it to their needs.

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

Designing for Situated Data

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

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

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

Conclusion

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Randall, Dave, Markus Rohde, Kjeld Schmidt, and Volker Wulf, eds. 2018. "Introduction: Socio-Informatics— Practice Makes Perfect?" In *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts.* Oxford: Oxford University Press. Accessed September 8, 2021. https://doi.org/10.1093/oso/9780198733249.003.0001.

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

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

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

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

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

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

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

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

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

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

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

Yuksel, Pelin, Bernard Robin, and Sara McNeil. 2011. "Educational Uses of Digital Storytelling All around the World." International Conference Society for Information Technology & Teacher, March 07, 2011, Nashville, Tennessee,1264–71. Waynesville, NC: Association for the Advancement of Computing in Education. Accessed September 8, 2021. https://www.learntechlib.org/primary/p/36461/.

IV: Datafied Mobilities

Mediating Affective Atmospheres through Public Wifi Infrastructure

Nathaniel O'Grady

Amongst the increasingly complex rubric used to articulate the imbrication of digital media in contemporary culture, the idea of a praxeology of data is one that, for me at least, is particularly compelling. When I was first told about the scope of this edited collection, the notion of praxis immediately evoked two overlapping points of reference. On the one hand, it encapsulated those actions whose accumulated re-occurrence proves constitutive of the routines and rituals that collectively shape and actively remake the shifting contours of lived reality in which data-based devices are evermore anchored. But on the other, it steered my attention towards the knowledges, logics and forms of sense-making enrolled into and emanating from the sites where digital media seep into these routines; allowing for rumination on how such media are generative of new ontologies of everyday life.

In this chapter, I read the infusion of data-based technologies into this two-headed rendition of praxis by tracing its expression through, and imprint upon, affective life and the emergence of atmospheres that are coextensive to these affects. Whilst affect refers to embodied forms of responsivity, and perhaps consequently some liminal form of sense-making, that are in continual negotiation between bodies and the wider ecologies they inhabit, atmospheres address how such affects hold in the air: existing in tension as collective feelings, vibes or moods that reverberate through space. Our complicity in such atmospheres reflects our capacity to affect and be affected. We actively contribute to atmospheres, but only through the potential that they unlock and activate for us. Contingent upon the myriad relations that make them up, affects and atmospheres are fluid; being continually made and remade amidst the flux of life.

Bearing ramifications for debates pertinent especially to new media studies, affects and atmospheres are, as Ben Anderson writes, "always already mediated" (Anderson 2014,13). During their invocation, affects reveal their inscription with the trace of the happenings that have preceded them. Affects are shaped by history whilst simultaneously operating as vessels for the sublimation of history into a present. Important for me in this chapter is that, if affects are said to express mediation, we see the heterogeneity of things that act as media too. Rather than solely referring to data-based devices, a more miscellaneous array of things will mediate the provocation of collective embodied moments in space. Discarded rubbish in a park, sticky floors on a train platform, the smell of baking bread, steam rushing forth from extraction pipes and an endless litany of other things might play their part in mediating affective experiences of space. And in a book about data and praxis, this intersectional and de-centered understanding of mediation allows us to consider digital devices not as isolated entities that mediate. Instead it focuses our attention on how such devices work in concert with a vast array of other things to figure within and actively help to modulate affective atmospheres that, albeit fleetingly, infiltrate, punctuate and surround the space-times of collective life.

I want to understand the political stakes where mediation is so defined, inquiring after how the curation of affective atmospheres via data practices is mobilized within, and indicative of, emergent forms of governance. To be more specific, I want to focus my attention on how digital devices penetrate situations and cultivate new practices therein that are based on an array of sensorial compulsions in pursuit of certain ends, guided by certain interests. Such technologies appear here as agents in what Brian Massumi has evoked as an "ecology of powers" (Massumi 2009, 173) that allows us to see the mediation of broader atmospheres via data practices as executed through an apparatus of power that seeks to "blend in with chaotic backgrounds" (2009, 153) where it operates. Across ecologies, the exercise of power is diffuse, taking place amongst an inchoate backdrop rather than being center-stage. Here is a modality of power that nudges across myriad sites both moving and static rather than dictating from a centralized position. This blending and dispersal, however, affords ecological power a temporality too, one in which those governing "must remain operationally open to unknowns and catch non-linear transversal phenomena" (2009, 154), seeking to adapt to the indeterminate swirl of the bodies and moods it vibrates through.

Whilst providing important precedents for considering how data practices furrow into atmospheres, recoding their affective intensity, an ecology of powers instigates careful re-appraisal of how perception, a key facet of affective life, figures in data practices oriented towards political ends. Perception addresses a whole manner of faculties used to sense surroundings. In some accounts, such sense-making is understood to be enacted on the basis of a cut that dichotomizes experience into binaries; of the imperceptible from the perceptible, say, or the visible from the invisible. But in this chapter I want instead to follow the path laid by Jenny Edkins in her traversal of such binaries (2019). An example of such a moment, according to Edkins, is the tension stoked between absence and presence in cases of missing people. Even though physical bodies may be categorized and treated *in absentia*, in other ways that person continues to imprint on spaces through sites they inhabited, objects that represent them or the memories of others. Even in their supposed absence, then, people remain present. Expanding this line of thinking and in contrast to some accounts (van Es and De Langa 2020), the vocabulary of affect and atmospheres enables a rethinking of the materiality of data in relation to the lived experience of urban life. The handprint that data practices leave in cultivating and modulating atmospheres do not conform to the language of perceptible/imperceptible, visible/invisible, absence/presence and so forth. Data-mediated atmospheres are, from the perspective of our bodily immersion within them, experienced on a precipice; constantly oscillating and undulating in their variance of intensity. Atmospheres will at one moment be strongly felt and, at another, hum lightly in the background. But never are they entirely evacuated from our collective sensorium.[1]

LinkNYC as Wifi Atmospherics

These conceptual reflections were provoked through research into LinkNYC: a public wifi infrastructure that has taken root and grown throughout New York City in recent years. An attempt to deliver on Mayor Bill De Blasio's 2012 pledge to widen access to the internet, the infrastructure is being developed through ongoing coordination between several public offices in New York and a conglomerate of for-profit companies that have taken CityBridge as their collective moniker for the project. The infrastructure supplies millions of denizens and tourists alike with a wifi connection that does not cost cash money, usually through their personal smartphones. In exchange, various sorts of data, outlined later, are extracted from users.

Like all infrastructure, Link is an assemblage composed of a litany of different agential materialities that entangle with one another in complex, and at times unpredictable, ways; serving different purposes. And with time these material forces have enveloped further still into the broader myriad constituents of urban milieus. The most visible manifestation of Link comes in the form of 10-foot kiosks that, in some parts of the city, stand at 150-metre intervals from one another up, down and across city blocks (see figure 1). These kiosks contain a tablet on which a limited number of web-based services can be accessed immediately. On either side of these kiosks are 55-inch screens displaying a catalogue of ever-changing adverts and a

1 Perhaps such an endless perceptivity is implied more widely by the technological changes that have accrued in recent years. In the smart city, reams of Big Data are constantly produced through a network of devices collectively paving the way for computing that is ubiquitous and cognition that is distributed. These are phenomena from which it is increasingly difficult to disengage. Data practices thus open up to consideration how digital infrastructures infiltrate and modulate affective atmospheres, steering their excess and ethereality towards certain ends. In so doing, a mode of governance is enacted that is diffuse, that moves through the spaces that atmospheres shroud and that possesses a certain dynamic: coming in and out of different levels of perceptive intensity.

few public announcements, related for instance to the dates of elections, community boards or emergencies (O'Grady 2021). But the further one furrows into its operation, the more diverse its components appear. It relies on an abundance of fiber optic cables running throughout the city. And of course, it incorporates human bodies, and the smartphones through which they connect to the internet, into its daily life too.

How, then, might that work outlined on affect and atmospheres, along with its reconceptualization of mediation and consequent effects for the perceptibility of data practices, make sense of the life of this digital infrastructure? Extending this question to its effects, what ramifications does it bear for our thinking about the fraught junctures between data practices, the governance of space and the movements therein? Below, I explore the repertoire of data practices that emanate where LinkNYC intersects with the broader corpus of people and things continually reshaping city life. Following these practices, I stay with the means through which LinkNYC infiltrates experience in and of the city whilst making this experience anew. As stated in strategic documents, it is by taking these experiences as its target that Link designers seek to make the infrastructure a so-called "native" element of the metropolis' quotidian. The chapter first expands on what it means for Link to strive towards nativity across urban scenes and the practices inaugurated in an attempt to reach this goal. In turn, I show how these practices invoke different affective responses at discrete points within the multitude of flows taking place through the city. These responses accumulate, becoming emblematic of the cultivation of new atmospheric conditions for urban experience, as mediated by data practices and all they encounter.

Figure 1: Example of a Link kiosk

Source: author's picture

Going Native

By insisting on atmospheres as space-times that undulate through changing registers of perceptive intensity, I follow ongoing reappraisals of the relationship between perception and the processes of individuation that always accompany the exercise of governance (see Simondon 1992). The act of perceiving represents a juncture at which individuals are carved out of, and gain some degree of autonomy from, a background situation to which they are bound; perhaps allowing a sense of self to arise in alignment with a broader set of power-relations. Massumi expands on such a process through elaborating on the notion of affective attunement (Massumi 2015). Here governance is inscribed in the processes through which, however gradually, humans begin to perceive themselves and the myriad things in their circumference as separate and distinct. Perception involves the exercise of various embodied capacities to distinguish objects from one another. And as these processes actualize evermore in the throes of the everyday, one becomes further and further integrated into homogenized modes of affective regulation. Surely, we are not yet so far beyond modernist ontologies that the idea of perceptual individuation, and its linkages to sovereign subject-hood, have entirely collapsed. Nevertheless, it does seem to be on its way out and I want to give it a little bit more of a push by suspending, perhaps temporarily, the notion of the imperceptible. In its wake, emergent practices of governance, that operate beyond renditions of the individual, can be said to incubate in the cultivation of atmospheres premised on fields of diffuse, ever fluctuating movement through moments of shared perception.

This invocation of post-individual spatiality and subjectivity is central to theories of atmosphere and affect more generally. We could, for instance, return to Anderson's unpacking of atmosphere and his conceptualization of them as irreducible to the sum of their parts. Atmospheres figure here as embodied moods that exceed and produce something novel beyond the relations that form the conditions of their possibility in the first place. They obfuscate archaic boundaries, blurring "the line between individual and collective" (Anderson 2014, 105). In his other work on the topic, Anderson extends this line of thinking, arguing that atmospheres render indistinct segregations between phenomena usually treated as oppositional, such that "to attend to affective atmospheres is to learn to be affected by the ambiguities of affect/emotion, by that which is determinate and indeterminate, present and absent, singular and vague" (Anderson 2009, 80).

Erin Manning has elaborated on what such a conflation of old dichotomies implies for perception specifically in her development of the idea of "autistic perception" (Manning 2016, 14). Adopting a post-individual vantage point to encounter the reality in which one is immersed, autistic perception "creates ecologies before they coalesce into form" (2016, 14). Environments are beheld here as an indiscriminate intermixture of things altogether in which the succession to archaic categorization is deferred, with "as yet no hierarchical differentiation" (2016, 14), for instance, "between colour, sound, light, between human and non-human, between what connects to the body and what connects to the world" (2016, 14). Bodies do not only simply sense, though. Instead they actively help to create that ecology via their sensorial responsiveness. Perception here does not individuate but is shared and affected across space. And if this is the case, all that mediates atmospheres, all that contributes to its continual remaking, brings with it a continual capacity to be felt at least on some register. This capacity to be felt might at times be latent but it nevertheless bears potential and, as such, weighs upon and reflects a situation by framing its virtuality; being emblematic of a possible trajectory for a future state of affairs that might arise and its conditioning in the present.

I want to suggest that LinkNYC operates within such a rendition of environments and post-individual subjectivity when it is deployed to make itself "native" within urban milieus. By making itself native the infrastructure looks to embed itself both materially and experientially in the city. Sometimes it will take the foreground in urban encounters and at others it will furrow into the background. But in either case it seeks to mediate and influence experiences in different ways. Inculcating itself as native involves (at least) two complimentary and interrelated practices. On the one hand, Link finds ways to enter and infiltrate pre-existing atmospheres. On the other, it also must continually readjust its function to the broader atmospheric flows it seeks to act in concert with. And these practices align to and actualize different registers of perceptibility of a broader atmosphere that Link mediates. For the rest of this section, I want to elaborate on each of these practices in turn.

In the first practice, LinkNYC is bound up in operations similar to those expanded on in literature concerning the extension of so-called ambient media through daily life. For Paul Roquet (2016) ambient media operates as a pin to orchestrate relations between heterogeneous, unruly things to actively cultivate a mood. In other words, it might take pre-existent spatialized things, constitutive of already existing atmospheres, and mediate by rearranging these things to create that atmosphere anew. For example, then, New York and its abundant, lively, changing atmospheres existed before Link appeared on the scene. To become native, Link infiltrates itself subtly into these atmospheres and adds to them – necessarily re-tempering their effervescence in the process. According to Intersection, the company behind the advertising campaigns executed through Link, this infiltration has occurred to the extent that they can claim that the infrastructure is now "part of the urban experience, offering media products that natively weave into people's lives as they journey through public space" (Intersection 2017).

But, and to come to the second practice, the dynamic of atmospheric regulation that going native accounts for expands further. Atmospheres, as I have already claimed, are constantly changing, being reaffected by the introduction of new things and how they ripple through space. And in its becoming native, Link's relationship within atmospheres must address this turbulence. Here the nativity that Link seeks draws semblance with the natality evoked in Hannah Arendt's work. For Arendt, natality addresses the miraculous creativity, or potential for such, that streams forth with the injection of every new human into the world. Arendt elaborates: "The miracle that saves the world, the realm of human affairs, from its normal, natural ruin is ultimately the fact of natality… the birth of new men and the new beginning, the action they are capable of by virtue of being born" (Arendt 1958, 247). Though natality is felt as it reverberates as a novel force, it also bears connections to historically entrenched conditions and processes. Labor and work, for instance, are for Arendt "rooted in natality in so far as they have the task to provide and preserve the world for, to foresee and reckon with, the constant influx of newcomers who come into the world of strangers" (1958, 9); newcomers that "possess the capacity of beginning something anew" (1958, 9). Theorists of affect share this interest with Arendt through their detailing of capacities (Anderson 2014), virtuality (Massumi 2002) and transversality (Deleuze and Guattari 1983). At a finer layer of resolution, all these notions branch away from one another, to be sure. But they nevertheless accentuate that working with affect means to be cognizant of the ever-present potential for change that constantly brims in our encounters in the wider more-than-human milieus we inhabit. The potential interwoven into affects can be extended to the atmospheres they invoke and are shaped by. If affects embody a proclivity to change then they exceed the present situation in which they are performed and felt. Constructed through affects, atmospheres possess a dynamism that reflects their turbulence. "Atmospheres," to return to Anderson, "are always forming and deforming, appearing and disappearing. They are never still, static or at rest" (Anderson 2014, 141).

Natality, along with these later concepts, highlight the excessive character of affect. These are compulsions-cum-feelings that, to be sure, are defined by their very tendency to escape capture. But this does not mean that the force of novelty is not aligned in ways, or indeed mobilized by, practices of governance (O'Grady 2019). For her part, Arendt depicts the bursting forth of natality as something not wholly controllable by the people from whom it supposedly derives. Putting something new into the world, by necessity, means that that thing is externalized and thus shared with, and in part appropriated by, the environments or atmospheres into which it infuses. Arendt sees this loss of proprietorship where natal actions are not claimed through their naming via language. "Speechless action" (Arendt 1958, 179), then, "would no longer be action because there would no longer be an actor, and the actor, the doer of deeds, is possible only if he is at the same time the speaker of words" (1958, 179).

Pre-linguistic novel affects may be enrolled into modes of governance through techniques that witness that affect's occurrence and translate it into the realm of the representative. In so doing, affects rematerialize as operable devices mobilized into processes directed at modulating collective atmospheres. What, I think, we are led to ask here is a quite simple question: what techniques can thus mobilize affects? To find a response, we might look to the forms of data capture LinkNYC deploys to understand its users and how these data are used recursively to change the operation of the infrastructure within urban environments; de facto producing affects anew and remediating urban atmospheres.

These practices of data capture course through different phases. And in different phases, the perceptibility of Link's presence in amongst urban atmospheres ranges in and out of different spheres of intensity for users. Firstly, data are collected about the users of the Link wifi network. After first connection, users no longer have to give Link permission to gather data from whatever device is connected. Over time, then, the collection process continues in a way that is increasingly interwoven as a taken-for-granted aspect of normal routine, becoming evermore surreptitious across the days, months, years that people connect to the network. A range of technical data are collected, including the MAC addresses of devices, the type of device, the language used by the device and the times between which the connection is sustained to the network (Intersection 2017). This real-time metadata are synced and integrated with wider, open source structured geo-demographic data-sets that estimate how many people will walk past and dwell amidst Link networks and their supposed attributes (Intersection 2017).

From the outset, Citybridge have been at pains to state that all the data they collect are anonymous. But anonymity is not interchangeable with the inability to identify. Nor is it necessarily against their interests to collect anonymized data. Striking a chord with the trans-individual character of affects and atmospheres, data at an individual level would not be that useful to the companies' strategy, whose primary concern is to understand and aggregate the scenes of collective life into which they wish to "become native." Through data collection, companies are able to infer much about network users. It is well documented that MAC addresses and device type data are used to interpret the level of income a person possesses and, perhaps more importantly, what they are willing to spend. Aggregated data on the times at which people connect to the network, when they are mobile and their destinations are considered integral elements to building up character profiles for users. The advertising company behind Link confirms as much, claiming that "we know people are always on the go and that their origins and destinations are strong indicators of who they are" (Intersection 2017). Addressing potential clients for Link advertising, the firm goes on to outline how it:

> takes a data-driven approach, working with… audience data sets to understand the daily journeys of your desired customers and target prospects. Based on that data, we then identify the highest value products for you for an unparalleled suite of urban media (i.e. Link kiosks) to reach exactly the customers you want, at times and places they are most receptive to your message (Intersection 2017).

It is the word "receptive" here that draws Link techniques to the natality Arendt describes and how specifically the infrastructure re-informs the atmospheres that it inhabits. From the data collected, what is called the Gross Rating Point of different people is inferred that establishes where and at what times they are likely to engage with Link screens. On this basis, the companies behind Link identify and seek to work on the potential of users by engaging with them at the specific times and in the specific spaces whereupon they are deemed most open to new forms of encounter; to absorb information projected onto screens and embrace with vim the possibility of new experiences and ways of life that teem within a range of products

– from phones to sportswear, drinks, and watches. With the arrival at adverts on screens, the data practices through which Link enters into and modulates atmospheres have registered at different spheres of perceptive intensity with users. Far from the surreptitious relation forged on the go where data are collected, with its advertising Link's presence is felt very prominently.

Bodies, Spaces and Shifting Perceptual Capacities

Perhaps implicated by these two practices I have described, bodies play a substantial role as sites through which Link pursues its goal of nativity. The embroilment of bodies here forces a reappraisal of how causality might be conceptualized amidst affectively charged atmospheres. Much of the time, affects are treated as embodied states whose arousal acts to reveal processes that have caused them. Affects are a sublimated reflection of some other process that bears less legibility than bodies and what they do. But affects are also revealed to be causal in themselves, getting caught up in mediating the generation of new shared moods. With Link, where atmospheres are tampered with amongst different registers of perceptibility, the causality riven through affects is different still. Affects might arise from the practices by which new atmospheres are brought about. But the practices in which they are enrolled are not registered that forcibly on the bodies that perform them. Causal processes run through, and act to reproduce, affects but perhaps at a lower, more surreptitious, level than has been written of before. Causal processes still take place through and upon the scenes of collective, embodied life but not necessarily to the extent of being registered cognitively. They exist prior to the disruptive violence of thought as Deleuze would have it (Deleuze 2004), whilst nevertheless instigating some perturbation in affects and the atmospheres that are co-extensive.

As described, affects are transindividual forces; arising amidst encounters between things. But being contingent on such relations, the intensity and significance of our embodied responses to space is continually in a state of flux; changing and shifting as new points of intersection are forged. Link kiosks have the capacity to capture data that reflects these encounters. Sensors can capture environmental data including matters of humidity, air pressure and temperature alongside air-pollutant data. Other data the infrastructure might collect relates to vehicles passing and sound levels. Through this data, Link can establish deeper, more intimate connections with the city and the bodies therein whilst remaining only at the edge of perception. What might be described as low-intensity encounters emerge here that carry on without stifling other ongoing engagements constitutive of daily life. So, for example, on my first trip to research LinkNYC in 2018, I immediately agreed to all the terms and conditions stipulated to connect to the network. On subsequent days I moved through the city; using the subway, my feet, buses and taxis. These trips were made at different times and for a variety of purposes. I travelled between Manhattan and Brooklyn for meetings in offices, to write up research notes, to eat and drink in bars. I exercised in parks. At all times my phone was on my body. And whilst I didn't interact with it, my phone was connected to the network without triggering my awareness. For large swathes of time my bodily, affect-laden, ever-changing encounter with the city was constantly recorded via the connection with Link, but Link's presence didn't stir any new feelings. Link might attune to what Erin Manning calls minor gestures here; habits and dispositions provoked through our ongoing response to the world that are usually taken for granted and not necessarily problematized. For Manning, however, these gestures are crucial for understanding how bodies become embroiled in the recreation of affectively charged space: "it is the minoritarian tendencies," she claims, "that initiate the subtle shifts that created the conditions for…any change" (Manning 2016,1) that takes place. In their activation, minor gestures embody the ongoing responsiveness of humans to the world; a responsiveness that shifts, no matter how subtly, the atmospheres in which bodies are enveloped.

These low-intensity encounters sit on a continuum alongside moments in which Link's presence is more prominent amidst our ongoing affective responsiveness to the city. Extending thinking on post-phenomenological theory by synthesizing aspects of speculative realism with constructivist approaches, James Ash offers the term inter-comprehension to consider the practices by which "entities relate to each other" (Ash 2020, 182). It is through these relations that material realities and co-extensive ontologies arise. But, says Ash, inter-comprehension is always guided by power relations, meaning that the enactment and ramifications for bonds forged are always distributed asymmetrically across the entities conscripted. Inter-comprehension is thus "actively designed to provoke, guide or otherwise influence the action or capacities of other entities" (2020, 187).

Coming back to Link, these processes of inter-comprehension appear as a designed practice in the calculations made to establish what marketing companies call impressions data. Rather than looking to target specific individual people through data capture, impressions data seeks to render legible the distance at which people's engagement with Link screens will be most intense and overwhelming for their perceptual capacity. Such an optimal distance is derived from the integration of various data that reflects on experiences within environments immediate to Link kiosks, including the type of street on which they are implanted, the size of the adverts they display, the speed of movement and dwell time around kiosks at different times of the day. Such data are synched with "Census population figures, Census population projections, the National Household Transportation Survey and the American Commuting Survey" (Geopath 2017). Once this intelligence is generated, strategization and decision making takes place to think through how adverts might be designed to take advantage of the maximum noting distance – thus seeking to modulate people's affective encounter with the city around them, orienting them to afford attention to adverts on screens.

Mobility's Collective Sensorium

Link's encounter with bodies is one that fluctuates, constantly oscillating across different registers of perceptual intensity. Perhaps this emphasis on undulating affects reflects a broader underpinning ontological assumption inscribed in Link's strategy to become native. This assumption is that perception morphs as bodies move. The encounters that Link seeks to instantiate conceive of perceptual capacities as they shift in motion. This connection between affect and movement is well established in literature. Such is elaborated by David Bissell in his book *Transit Life* where he argues that, caught up in movement, bodies express their openness to new forms of experience and their capacity to act in new ways as they continuously engage with the world. This dynamism that bodies-in-movement evidence paves the way for further conceptual extension to fathom the connections between bodies and what Bissell calls "ecologies" (Bissell 2018, 163). Figured as a "complex web of relations with other people, places, times, ideas and materials" (2018, xix), ecologies leave a trace on bodies, shaping how they feel, move, and make sense of their surroundings. Amongst an ecology's affectivities, however, bodies impress themselves upon the spaces they inhabit too. These ongoing co-constitutive negotiations reveal that bodies are ever-embroiled in processes of "enablement and constraint" (2018, xxi). By considering how bodies can move and how they cannot, in other words, we can start to consider the ways in which environments figure in the mediation of movement via practices of governance (Adey 2008).

Such practices of enablement and constraint are certainly present when exploring how Link imbricates itself in the movements that contribute to the atmospheres in which it operates. However, we might add another layer to the terms used here in exploring the practices enacted by Link. In particular we might think of integration as a form of enablement and incorporation as a mode of constraint. Each of these practices bear upon movement across urban space through instantiating specific modalities of relation between bodies in motion and Link infrastructure. Regarding what I have termed integration, this practice is enlivened through modifications that have been made to Link infrastructure to ensure it is synchronized with the myriad flows that in part constitute life in New York. Link has been designed to syncopate with the polyphonic currents whose regularity accumulates to form the overarching waves that shroud the city at different points in the day. Such syncopation is evident when our gaze is drawn towards the ongoing development of the status of Link kiosk's tablets through which users access the internet. When the kiosks first appeared, users could spend unfettered amounts of time on the units.

But after numerous scandals involving the content being accessed on the tablets, the hours some would while away on them and the masses of people that sometimes would gather around, it was suddenly declared that "The LinkNYC tablet is meant to be an on-the-go resource" (Intersection 2017) and that, in a bid to "curb long term use of the kiosk" (2017), tablet interface use for a single session would be limited to 10 minutes and that internet services would be withdrawn after 1 minute of inactivity. Resembling a Foucaultian biopolitics of conducting conduct (Foucault 2008), perhaps the overwhelming rationale for these restrictions was to ensure that Link does not block the circulation of people and things that flow through Manhattan's bustling sidewalks. But the effect of this modulation is to suture and enfold Link infrastructure into routine flows, thus acclimatizing its existence into the rhythms of city life. Such was evermore evident when I discussed the material form of Link kiosks with their designers, who described how their shape, size and position was molded to facilitate egress that brings life to New York's streets. And the further it seeps into the urban milieu's background, the more perceptually normalized Link becomes.

Where movement, mobility and circulation take center stage in our analysis, nevertheless, Link also develops in the reverse direction in relation to its perceptibility. In other words, rather than reshaping its own functions to become part of city life, Link interweaves itself perceptibly into people's daily life through its gradual incorporation into their routine movement. Such a maneuver can be witnessed where Link establishes relations to the sensory capacities of people and orients them towards certain ends, thus showing its investment into what Bernard Stiegler calls a "retentional economy" wherein technologies are deployed to mediate consciousness (Stiegler 2010). Returning to his work, James Ash has drawn on the notion of retention to show how computer game design involves a "series of retentional ecologies and environments… to generate particular forms of affect" (Ash 2012, 7) that work "to capture and hold users' attention" (2012, 6). Link's targeting of retention though is different from that which Ash describes. For Ash, retention is something that will be captured and sustained for a length of time that is strung out. Eyes and bodies might be sealed to screens for hours. Retention is an object that Link seeks to hold, conversely, both more fleetingly whilst also being something it seeks to inculcate gradually over time. It is also something that Link seeks to harness on and off as people move in and out of places, through day and night.

Link's mobilization of retention is evident in the strategy articulated by the marketing company behind the adverts that flash up on Link screens. This strategy is first to understand the context in which movement takes place as ascertained through "data feeds including local weather, events, maps, traffic, social media and more" (Intersection 2017) to identify "the critical data consumers need to inform their journeys" (2017) and then present this data on kiosk screens. Link advertising here creates fleeting encounters between screens and people that are relevant to people as they move through the city. Over time, these micro-encounters become part and parcel of the normalized perceptual range of people's routine, rhythmic movement through the city. They peer occasionally at screens to attain useful information. However, much of the time when they look at the screens they will not see information on the chance of rain, say, or delays on the subway but adverts for trainers, perfume, holidays and so on. As the marketing company explain themselves: "The digitization of assets in transit authorities has created communication platforms that display emergency service announcements, provide real-time train updates and offer contextual messaging. This is retraining consumers to look at screens more actively, increasing the value of "adjacent advertising." Already, we are seeing brands natively weave themselves into the context of this messaging, providing utility as well as engagement" (2017). As it becomes an increasingly integral source for information on matters considered crucial for commuters, LinkNYC simultaneously infuses into urban atmospheres the interests of companies whose services and products are anything but.

Conclusion: Choreographing Affects for a Fluctuating World

In this chapter I have expanded on data practices by tracing their entrance into and modulation of the turbulent maelstrom of affects that arise from and cast themselves, however momentarily, across the space-times of urban life. Synchronized with flowing affects, the modes of mediation that data practices are enveloped into are wholly decentered from any point of technological interface with a specific, singular device. Mediation is instead a process continually negotiated amidst multiple heterogeneous material agencies out of whose infusion arise fleetingly coherent spatial ensembles that might be known by various names; whether ecologies, atmospheres, environments or milieus for example. Manifest by its inscription onto bodies and their capacities, the forms of mediation that data practices enact and become enrolled in follow a particular choreography. Bodies affect. Before any point of conscious revelation, they operate as vessels bringing the situation from which they have arisen to bear on other spaces, thus mediating new situations entirely. But at the same time, bodies are affected. Upon ever-shifting registers of perception, affects bear the imprint of the modes of mediation that bring them about and, in so doing, express their shaping through forms of governance. Extending Massumi's work, affects show that data practices help to undergird an ecology of powers that infiltrates and recursively nudges forms of encounter that characterize our experience of urban scenarios. Such is evident in this chapter by the proprioceptive repertoires of bodies across the city: from shifting eye trajectories to changing relations to smart phones that rest intimately on our bodies.

But in their affectation, such repertoires end up affirming and in turn actually modulating new forms of knowledge and rationalities that have developed to sustain these ecologies. Bodies thus speak to the very strategies that seek to suture them into environments molded by data practices. At the same time, it is important to remember that such knowledges are reflexive, they learn from and adapt to the modes of encounter that bodies register in these environments and the range of perceptive capacities actualized in tow. These knowledges mirror the emergence of new processes of subjectification that do away with the idea that an individuated subject can be identified and constituted as such by its inheritance of sensory detachment from wider environments. Affects are trans-individual – emanating from and rebounding through spaces prior to the setting in motion of forms of attunement through which people are taught to extricate themselves from environments via learnt modes of cognition. Considering bodily capacity as its target-object, governance that seeks in some way to address affects looks not towards individuals, then, but to the intersections between entities that encounter one another in and through space.

Building pathways to attend to the integration of data practices into affect-laden space-times and their mobilization as an object of governance presents some serious methodological challenges whose difficulties far exceed this chapter. But some of the questions that might be asked can at least be formulated here. How might we evoke the felt permutations that fluctuate as encounters are renegotiated through the presence of data practices? What efficacy would our commentary hold by attempting to bring into focus modes of mediation that are shared, diffuse, decentered and eminently "ecological"? Is it not folly to grasp for and represent experiences whose liveliness is constituted by the fact they exceed representation? Perhaps some promise to finding a response to these questions might lie in developing further Maria Puig de la Bellacasa's call for a poetics of infrastructure. Inspired by Susan Leigh Star's work, such a poetics involves making sense of practices that have become routine by engaging with what has been erased through their very stabilization, thus expressing "other possible worlds hidden or silenced in marginalised spaces" (Bellacasa 2016, 49). Whilst such a poetics orients our focus to things that exist beyond the scope of this paper, LinkNYC does show us that affects we may not even register might nevertheless be enrolled into data sourcing, meaning their silence is not ensured. Indeed these barely perceptible affects figure in practices crucial to the infrastructure's operation. So perhaps what poetics needs to be supplemented by is a sense of the ever-present existence of these alternative worlds and the potential they bear for redressing the implications that data practices bear upon everyday atmospheres in the city.

References

Adey, Peter. 2008. "Architectural Geographies of the Airport Balcony: Mobility, Sensation and the Theatre of Flight." *Geografiska Annaler: Series B, Human Geography* 90 (1): 29–47.

Anderson, Ben. 2014. *Encountering Affect: Capacities, Apparatuses, Conditions*. Farnham: Ashgate.

Anderson, Ben. 2009. "Affective atmospheres." *Emotion, Space and Society* 2 (2): 77–81.

Arendt, Hannah. 1958. *The Human Condition.* Chicago: University of Chicago Press.

Ash, James. 2012. "Attention, Videogames and the Retentional Economies of Affective Amplification." *Theory, Culture & Society* 29 (6): 3–26.

Ash, James. 2020. "Post-phenomenology and space: A Geography of Comprehension, Form and Power." *Transactions of the Institute of British Geographers* 45 (1): 181–193.

de la Bellacasa, Maria Puig. 2016. "Ecological Thinking, Material Spirituality, and the Poetics of Infrastructure." In *Boundary Objects and Beyond: Working with Leigh Star*, edited by Geoffrey C.Bowker, Stefan Timmermans, Adele E. Clarke, and Ellen Balka, 47–68. Cambridge, MA/London: MIT Press.

Bissell, David. 2018. *Transit Life: How Commuting is Transforming our Cities*. Cambridge, MA/London: MIT Press.

Crary, Jonathan. 2013. *Late Capitalism and the Ends of Sleep*. London: Verso Books.

Deleuze, Gilles, and Felix Guattari. 1983. *Anti-Oedipus: Capitalism and Schizophrenia*. London: A&C Black.

Deleuze, Gilles. 2004. *Logic of Sense*. London: Bloomsbury Publishing.

Edkins, Jenny. 2019. *Change and the Politics of Certainty*. Manchester: Manchester University Press.

Foucault, Michel, Arnold I. Davidson, and Graham Burchell. 2008. *The Birth of Biopolitics: Lectures at the Collège de France 1978-1979*. Wiesbaden: Springer.

Geopath. 2021. "Geopath Transit Research." Accessed May 3, 2021. http://geopath.org/wp-content/uploads/2017/12/GeopathTransit.pdf.

Geopath. "Geopath Glossary." Accessed January 28, 2021. https://geopath.org/glossary/#:~:text=who%20notice%20it.-,Gross%20Rating%20Points%20(GRPs),1%25%20of%20the%20market%20population.

Intersection. 2017. "Why tech is Going Traditional in Transit." Accessed January 28, 2021. https://ixn.intersection.com/why-tech-is-going-traditional-in-transit-7dda083d82d4.

Manning, Erin. 2016. *The Minor Gesture*. Durham: Duke University Press.

Massumi, Brian. 2002. *Parables for the Virtual: Movement, Affect, Sensation*. Durham: Duke University Press.

Massumi, Brian. 2009. "National Enterprise Emergency: Steps Toward an Ecology of Powers." *Theory, Culture & Society* 26 (6): 153–185.

Massumi, Brian. 2015. *Ontopower: War, Powers, and the State of Perception*. Durham: Duke University Press.

O'Grady, Nathaniel. 2019. "Designing Affect into Security: Shared Situational Awareness Protocols and the Habit of Emergency Response." *Cultural Geographies* 26 (4): 455–470.

O'Grady, Nathaniel. 2021. "Automating Security Infrastructures: Practices, Imaginaries, Politics." In *Security Dialogue*, 52 (3): 231–248.

Retail Customer Service. 2018. "Link NYC Kiosks Improving Quality of Life in the Big Apple." Accessed January 28, 2021. https://www.retailcustomerexperience.com/articles/linknyc-kiosks-improving-quality-of-life-in-the-big-apple.

Roquet, Paul. 2016. *Ambient Media: Japanese Atmospheres of Self.* Minnesota: University of Minnesota Press.

Simondon, Gilbert. 1992. "The Genesis of the Individual." In *Incorporations: Zone 6*, edited by Jonathan Crary, and Sanford Kwinter, 296–319. New York: Zone Books.

Stiegler, Bernard. 2010. *For a new critique of political economy*. London: Polity.

Dashboard Design and Driving Data(fication)[1]

Sam Hind

In this chapter I consider how the re-design of vehicle dashboards has restructured car-related data processes. I do so by charting the emergence of two such processes enabled by the re-design of vehicle dashboards. Firstly, the transformation of "geodata" into "navigational data" with the integration of voice-activated navigation systems into vehicle dashboards. Here, this transformation is enabled through the implementation of new addressing and speech protocols that radically change the relationship between driver and vehicle, when performing navigational tasks. Secondly, the transformation of "vehicle data" into "driving data" in the convergence, and customization, of dashboard features and functionality. Here, this transformation is enabled through the spatial, aesthetic, and operational integration of typically separate aspects of the driving experience (instrument cluster, navigation, entertainment), re-presenting vehicle-related data in new, and novel, ways. In evaluating these concomitant "datafication" processes, I use Mejias and Couldry's (2019, 3; emphasis added) definition, in which datafication involves "the transformation of human life into data through processes of *quantification*, and the generation of different kinds of *value* from data."

Both transformations are enabled through strategic design decisions, persuading drivers to participate in novel practices they might otherwise not. Firstly, through the strategy of "representational transparency" (Agre 1995, 186), in which voice-activation is depicted as a seamless, unmediated interface (Bolter and Grusin 2000) between the normal, natural speech of a driver, and the vehicle itself. Secondly, through the strategy of control, in which the driver is persuaded to believe they have full(er) customizable power within, and critically *of*, the vehicle – an example of what Mattern (2015, n.p.) refers to as "dashboard drama," or an aesthetic allure in which the driver is "empowered" (Agre 1995, 175) through the customization of their vehicle, that also results in their driving experience being managed by the vehicle manufacturer.

1 This chapter was first published in expanded form as Hind, Sam. "Dashboard Design and the 'Datafied' Driving Experience." Big Data & Society (volume 8, issue 2, July 2021) under Creative Commons license CC-BY-4.0.

Here, I interrogate how such systems transform car-related data from one state (geodata, vehicle data) into another (navigational data, driving data). The systems discussed here are representative of broader efforts within the automobile industry to transform the vehicle itself into "mobile spatial media" (Alvarez León 2019a) or wholesale into a "platform," through which the use of data is integral (Wilken and Thomas 2019; Alvarez León 2019b).

Whilst the automobile industry is not alone in making use of data streams produced as by-products (Thatcher 2014; Pridmore and Mols 2020), there are nonetheless unique challenges to be found in this application, such as interpreting spoken destinations or disambiguating common street names. These provide the possibility of articulating distinct aspects of datafication (van Dijk 2014; Sadowski 2019) within vehicles, and beyond other spaces such as the home (Pridmore et al. 2019; Maalsen and Sadowski 2019). The effect is manifold: the cultivation of new kinds or streams of data (touchscreen interfaces augmented with voice-activation, mirrors replaced by recordable cameras), new examples of representing established kinds of data (vehicle speed, or fuel levels), and altered practices in relation to both (entering destinations, checking mirrors).

The aim of the chapter is thus threefold. Firstly, to map where and how datafication takes place within the car. Secondly, to establish the role of vehicle dashboards in enabling this datafication. Then thirdly, to identify the strategies that come to shape the nascent "datafied" driving experience.

In the next two sections I consider how geodata is (and is not) transformed into navigational data, and how vehicle data is transformed into driving data. In the former, I discuss how some kinds of geographical information escape datafication, whilst others are subject to a practice I call "re-datafication." In the latter, I discuss how vehicle data is "surfaced" as driving data, generating alternative kinds of value in the process (Mejias and Couldry 2019). In the subsequent section I explore how dashboard "convergence" (Jenkins 2006; Hind and Gekker 2019) enables these transformations. In the final two sections I discuss two cases: a voice-activated navigation system built on the What3words platform, and a "widescreen" dashboard in a range of Mercedes-Benz vehicles.

Navigational Data: "Turn-by-Turn"

Geodata is data with a geographical, locational, or spatial component (Lauriault 2017; Leszczynski 2017). Typical examples include coordinates, a house address, or a postal code. Geodata may be relatively precise (GPS location) or represent a general geographic area (a state, or municipality), with scholars attentive to the spatialities of data, more broadly (Crampton et al. 2013; Shelton 2017). Further, geodata can be used for various kinds of navigational tasks: to orientate oneself on a hike, to enable the delivery of consumer goods, or to arrange a meeting with friends. In this section I consider first how some kinds of geographical information are *not* transformed into data in the act of navigating a vehicle. Then, how geodata *stays* as geodata whilst being enrolled into navigational practice, before discussing how geodata *is* transformed into navigational data in the act of navigating a vehicle. In other words, how navigational data is "activated" through various associated practices, which to varying degrees satisfies a general definition of datafication, as offered by Mejias and Couldry (2019) involving both (a) quantification and (b) generation of different kinds of value.

Navigational data is dependent on geodata. Geodata might be *added* to other kinds of (geo)data such that its use, or operational, value is enhanced. Geodata might also be *replaced* by more useful geodata that usurps the original geodata's low utility. Both enable a navigational task to be completed. Thus, navigational data is always composed out of geodata but may be combined with other contextual data that aids the completion of the navigational task. If there is no navigational task to perform, the geodata remains as geodata.[2] Even then, geodata may be enrolled into navigation without being transformed into navigational data. In any case, navigational data does not exist *a priori* but is transformed into navigational data through the act of navigation. It is, therefore, "ontogenetic," emergent in the practice of navigating (Kitchin and Dodge 2007; Hind 2020).

Navigational data is *not* always activated in the driving of a car. Firstly, geographical information may be embedded within the wider environment, both inside and outside the car, that remains as geographical information but is still integral to navigational practice. This may include visible buildings or landmarks and temporary road signs that issue text-based instructions, but also trusty road atlases or "occasion maps" scribbled on rough pieces of paper (Singh et al. 2019; Thielmann 2019). Although there may well be a change in value as these phenomena are enrolled into specific navigational tasks, they are not turned into (geo)data in the process, and as such do not undergo datafication.

Secondly, geodata may become enrolled into navigational practices, but stubbornly remain as geodata. Typically, such geodata might be found in fixed road signs with place names and numerical distances, or in traffic lights, in which geographical information has already been extracted, coded, and displayed. Yet, whilst enrolled into navigational practices (think of how many times a motorway road sign has been interpreted by a passing driver), this geodata is not transformed into navigational data, as no further quantification has taken place, even though different kinds of value are arguably being generated through its enrolment in many spe-

2 Whilst it is beyond the scope of this chapter, it is also possible for geodata to be transformed into other kinds of data, besides navigational data.

cific, unique navigational tasks each day. Thus, it cannot be said that additional datafication happens in such a case.

Thirdly, geodata may be enrolled into navigational practices that transform it into navigational data. Ordinarily, this activation occurs through digital devices such as integrated sat-navs, standalone sat-navs, generic or single-purpose map apps (Brown and Laurier 2012; Chesher 2012; Hind and Gekker 2014; Hind 2019). With each, geodata undergoes a second round of datafication, as pre-existing datapoints (house numbers, postcodes etc.) are transformed into "turn-by-turn" datapoints in the act of navigating via a digital device, or what Singh et al. (2019, 287) refer to as a "turn-taking machine," thus generating immediate navigational value to the driver. Rather than datafication, per se, this "re-datafication" instead further transforms one kind of data into another. These transformations are necessarily performed in part (or whole) through specific technical relations between devices, apps, platforms, and infrastructures, that differ between the examples given above.[3]

This final category is of particular interest because of how it usurps these other modes. When geodata is transformed into navigational data it captures and codifies the navigational experience, "turn-by-turn." However, these other sources of navigational information are also rather stubborn: they stand in the way – sometimes quite literally – of datafication, limiting the extent to which users might require, or interact with, navigational devices. It is this contestation between reliable, appropriate, and accurate sources of navigational information that I will turn to in the first case study.

In summary, whilst in the first category (geographical information in navigation) value generation might occur, this is not through datafication. In the second category (geodata as geodata) datafication has already occurred, and value generation does take place, but not through the transformation of geodata into navigational data. In the final category (geodata into navigational data), "re-datafication" can be said to occur, through which both quantification and value generation take place in the act of navigation.

Driving Data: "Smarter Decisions"

Vehicle data is data that is generated by the car for the technical operation of the vehicle itself. Ordinarily, the transmission of vehicle data is enabled through a centralized communication system, referred to as a "vehicle bus." Various protocols

3 For instance, in how the capacities of map apps lend themselves to more seamless and continuous modes of datafication than integrated sat-navs.

have been developed over the years to standardize these communication procedures, ensuring components can safely and effectively speak to each other. "Electronic control units" (ECUs) that control a suite of functions within the car are dependent on vehicle data, as well as the vehicle bus that sends such data throughout the vehicle. Examples include the engine control unit (for controlling the engine), the transmission control unit (for controlling gear transmission), or the control unit for anti-lock braking systems. Luxury cars now have between 100 and 150 ECUs (Stoltzfus 2017; Winning 2019), despite attempts to consolidate them into multifunctional systems (Intel 2018).

As vehicle data facilitates machine-to-machine communication, it does not need to be seen by, or made interpretable for, the driver. However, much like geodata is transformed into navigational data, so vehicle data can be transformed into driving data, surfaced through representation in indicators, dials, lights or some other visual (or audio) form. Driving data is vehicle data that is activated in the process of driving a vehicle. Whilst vehicle data lies under the bonnet, (usually) quietly ensuring the vehicle is operating properly, driving data is presented to the driver to aid decision-making. Like navigational data, driving data is brought into being through the various stages or moments in the driving experience.

Vehicle data is surfaced as driving data in multiple ways. Firstly, data can be *proximally* surfaced. Whilst the manual use of external indicators and headlights (to express "thanks") (Brown and Laurier 2017) are examples of proximal communication, these do not typically require ECUs, and thus do not generate data at all. As visual signs, they are analogous to the geographical information category discussed previously. Whilst such instances may well generate value, as these phenomena are enrolled into specific driving actions, they are not turned into (driving) data in the process, and as such do not undergo datafication. However, the development of "adaptive" or "intelligent" sensor-activated headlights that dynamically adjust to different conditions (fog, night) or situations (an urban environment, a sharp corner) do require ECUs and as such undergo datafication. Here, driving data is not only made available for other road users to aid safety and ensure appropriate driving etiquette, but is also dynamically enrolled into driver decision-making.

Secondly, such data can be *internally* surfaced. Here, data is surfaced through instruments, dials, lights, and screens on a vehicle dashboard. Such data is principally surfaced for the driver, to ensure they can perform driving activities, such as deciding when to refuel or change gear. In the USA, 44 separate indicators are standardized by law (Federal Motor Vehicle Safety Standard 2020). In 2020, Honda recalled 608,000 vehicles in the USA (O'Kane 2020) after discovering faulty software that could "cause the instrument panel to not display critical information" (National Highway Traffic Safety Administration 2020, 1). As vehicle data has already undergone a process of datafication (quantification and value generation), the transformation into driving data can be considered as re-datafication, through which new safety-critical forms of value are generated.

Thirdly, data can be *remotely* surfaced. Here driving data is extracted for the use of remote parties such as car rental companies, haulage firms, or insurance providers (Meyers and van Hoyweghen 2020). "On-board diagnostics" (OBDs) are typically used to track vehicles, with simple devices plugged into OBDs – i.e. transforming vehicle data into driving data. For fleet operators, external surfacing enables "smarter decisions, powered by data" (US Fleet Tracking 2020, n.p.), through which vehicle assets can be managed. Increasingly, however, this type of extraction is being enabled to both obtain ever-more granular driving data, as well as to expand such efforts to everyday vehicle owners (Gekker and Hind 2019). In this case, vehicle data is transformed into driving data through a re-datafication process that yields greater opportunity for the aggregation, combination, and comparison (i.e. quantification) of, and between, such data. This results in a more intensive and persistent generation of value, mostly for the parties above, but also potentially for other drivers as insights gained from the re-datafication process inform the redesign of vehicle dashboards and associated technologies.

In summary, in the first category (proximal surfacing) datafication occurs under certain circumstances, with recurring, and familiar, forms of value generated between drivers in the process. In the second category (internal surfacing) datafication has also already occurred, with secondary re-datafication processes seeking to cultivate new forms of value, principally for drivers. In the final category (remote surfacing), re-datafication results in more aggressive forms of data use for third parties, through which a multitude of different forms of value might be generated. Some of these final forms of value may find their way back to the driver in the redesign of vehicle dashboards. The following section will thus consider the role of the dashboard in the datafication process.

Dashboard Convergence: From Spatial to Operational Integration

Originally a physical board to prevent material from "dashing up" onto the exposed driver (Mattern 2015), dashboards in contemporary vehicles display an array of phenomena from operation-critical processes (gear, engine temperature) to multimedia options (radio station, Bluetooth connectivity). In this, the dashboard has developed from a device made to prevent any physical hinderances to driving, to a multifunctional aid meant to enhance the driving experience.

In this section I contend that dashboard "convergence" facilitates datafication, in which the otherwise separate interfaces housed within a vehicle dashboard – typically the instrument cluster, navigational assistance, and multimedia – are becoming operationally dependent. In this, I go beyond Alvarez León's (2019a, 370) suggestion that cars have become "integrated media spaces," arguing that it is through this convergence that the datafication of wholesale vehicle operations is occurring.

Dashboard convergence is not necessarily a new phenomenon, with innovation in vehicle dashboard design being a constant since the early 20th century. Yet, as Mattern (2015, n.p.) discusses, the "standard package" of a Ford Model T in 1908 "consisted solely of an ammeter, an instrument that measured electrical current." Whilst early computers only made use of displays to check for "errors" rather than for "complex data output or input" (Thielmann 2018, 47), early motor vehicles had more immediate ways to inform users of a problem: "[w]ater gushing from the radiator, an indicator you hoped *not* to see, was your 'engine temperature warning system'." (Mattern 2015, n.p.; authors' emphasis). Yet dashboards had already became symbolic representations of vehicle state, rather than strictly indexical representations, "progressively simplify[ing] the information relayed to the driver, as much of the hard intellectual and physical labour of driving was now done by the car itself." (2015, n.p.).

Following Mattern, then, dashboard design from the 1950s onwards exhibits a kind of rationalizing design tendency, with fewer vehicle operations needing to be represented (either indexically or symbolically) to the driver. However, the dashboard has become an important space for innovation in recent years as new driving features and data flows are represented. Dashboard convergence is both a preparation for, and a logical effect of, the "platformization" of the car, made connectable, modular, and interface-able (Helmond 2015).

This convergence is two-fold. Firstly, there is a *spatial* convergence in which previously separate modules are placed within the same part of the vehicle dashboard. A typical example is the integration of navigational capabilities within a multimedia system embedded within the central console of a vehicle (Alvarez León 2019a). Here, an external navigational device (sat-nav, road atlas, occasion map etc.) is replaced with an in-built feature, selectable by the driver in the same way as they turn on the radio, or adjust the air conditioning. In this, multimedia and navigation functions exist on the same ontological plane – as "apps" – embedded within such a system, accessible via buttons or a touchscreen.

Secondly, there is *operational* convergence involving the integration of previously separate systems. In this form of convergence, different systems are "vertically" integrated, such that either one is dependent on the other. Arguably, this operational convergence is newer, a required step towards the platformization of the car, in which different systems are made interoperable, in a "plug-and-play" approach, similar to how web platforms offer access via APIs (Plantin et al. 2016). A novel example, to be discussed in the next section, is the integration of voice-activation systems with navigational capabilities. Here, a voice-activation system, a unique addressing system, as well as an infotainment system, work together, with commands issued through one (voice-activation), triggering a response in another (addressing), to be presented in another (infotainment).

These two types of convergence – spatial and operational – are integral to "re-datafication": the transformation of geodata and vehicle data into navigational data and driving data. This horizontal (spatial) and vertical (operational) integration enables the activation, surfacing and/or extraction of these data types, in ways that were either previously unimaginable, technically impossible, difficult to implement, or otherwise "siloed." Through re-datafication, navigational and driving data streams are made more valuable, both for the driver of the vehicle, and – just as critically – for the manufacturer. Through this "interoperability" (Wilmott 2016) previously separate systems (and their accompanying data types) are made to work with each other.

In the following two sections I analyze how the question of convergence has been addressed in relation to two innovative vehicle dashboard designs: firstly, a voice-activated navigation system based on the unique addressing platform What3words. This can be seen to have transformed geodata into new forms of navigational data. Secondly, a self-styled "widescreen cockpit" designed by Mercedes-Benz, which has arguably surfaced vehicle data as novel forms of driving data. The result of both is an emerging datafied driving experience. I contend that their comparative successes – as examples of dashboard convergence in which data is transformed within the vehicle – rest on two strategies: representational transparency, and customizable control.

Voice-Activated Navigation as Representational Transparency

What3words is a geocode system that divides the world into $3m^2$ grids, identifiable by a unique three-word string. In doing so, it converts underlying geographic coordinates, enabling users to remember locations such as "thread.strollers.bumble" (somewhere in Germany). What3words claim it is superior to established postal systems, with the $3m^2$ grids enabling users "to specify a precise entrance, unlike a street address which identifies an entire building" (Macgregor 2020a, n.p.), and that "unlike street addresses which are often duplicated," What3words locations are "all unique," available in "over 40 languages" (2020a, n.p.).

1: An illustration of how the What3word grid system works

Source: What3words

In 2018 Daimler integrated What3words into its new infotainment system, "Mercedes-Benz User Experience" (MBUX), aboard the new Mercedes-Benz A-Class (Daimler 2018a, n.p.). In doing so, they suggested they had "moved one big step closer to [their] goal of making the vehicle into a mobile assistant" such that "[i]nputting locations…makes life easier for our customers and ensures a special experience" (Daimler 2018a, n.p.). In this, What3words could be activated through a touchscreen interface, but also via voice control.

As a promotional video demonstrates (Mercedes-Benz 2018a),[4] the combination of What3words and LINGUATRONIC (MBUX's voice control system) (Daimler 2018b) is seen as integral to the navigational experience within the A-Class. Rather than the driver using clunky search boxes, unresponsive knobs and buttons, or external apps or sat-navs, the owner merely issues instructions to the vehicle, with What3words deemed "the simplest way to talk about location" (Mercedes-Benz 2018a, n.p.). In this, typical (local) addressing systems are rendered confusing and frustrating. As the manufacturer reminds us, street names are rarely unique (Mercedes-Benz 2018b, n.p.), street numbers are difficult to differentiate between, and some towns are entirely unpronounceable to unfamiliar users (What3words 2019, n.p.). In short, the integration of What3words into the A-Class is set up to provide an improved navigational experience.

4 The original video has since been removed, however the page is still accessible via an archive link. A similar example can be found here: https://what3words.tumblr.com/post/179683783794/all-you-need-to-know-about-mercedes-benz-gps-voice

Figure 2: The frustration of inputting addresses by touch

Source: What3words

However, rather than simplifying vehicle navigation, the convergence of a novel addressing system, voice control system, and navigational system yields significant "praxeological changes" (Thielmann 2018, 50) between driver and vehicle. This operational convergence demands drivers follow, very carefully, a set of new conversational protocols to activate the navigational experience. Rather than offering a "transparent" interface, in which the "illusion of representational transparency" (Agre 1995, 186) between the driver and vehicle is maintainable, What3words "remediates" (Bolter and Grusin 2000) this navigational experience, inserting a number of new rules drivers must follow in order to navigate.

As a promotional video shows, "Sophie" receives a message on her smartphone: "let"s meet at hello.page.brand for brunch" (Mercedes-Benz 2018a, n.p.). As she gets into the car, she utters the words "hey Mercedes," before asking the vehicle to "take [her] to What3words hello.page.brand" (Mercedes-Benz 2018a, n.p.). Yet rather than instantly generating a route, Sophie is instead given three choices: her intended destination, but also "hello.page.brands" as well as "hello.page.barn" (Mercedes-Benz 2018a, n.p.). Whilst What3words is designed to remove ambiguity, only a plural form of one word distinguishes two results (brand to brands). Further, that despite their claim that "What3words addresses are spaced as far apart as possible to avoid confusion" (Macgregor 2020b, n.p.), all three options are within a 23-mile (37km) radius of Sophie's current location. Likewise, upon changing plans, a second set of results yields three locations in barely a 30-mile (48km) radius. Again, the three-word addresses aren't readily distinguishable, with "lanes.larger.daring" returned alongside "lands.larger.daring" and "ages.larger.daring" (Mercedes-Benz 2018a, n.p.).

The activation of navigational data in the use of a voice-activated navigation system results in a peculiar, and novel, experience. Geodata in the form of a typical, localized address is rendered mute, to be replaced in search results with both a broader reference to a geographic area (Bayswater, Cranleigh or Send) and the What3words addresses using words "assigned by a mathematical algorithm" using "simpler and more commonly-used words in each language" placed "in the areas where the language is spoken" (Macgregor 2020b, n.p.). In other words: "hello.page.brand" rather than 59A Portobello Road, Notting Hill, London W11 3DB.

This erasure – of localized names and places, idiosyncratic, ambiguous or unpronounceable to outsiders – is common to (digital) capitalism (Rose-Redwood et al. 2019; Sotoudehnia 2018; Nicas 2018).[5] The automatic translation, by "mathematical algorithm," otherwise a "toponymic reconfiguration" (Rose-Redwood, Alderman, and Azaryahu 2010, 454), of established place names into arguably banal if not trivial three-word locations, is key to the reorganization of navigational practice within the vehicle. With this, the driver is expected to change their navigational habits on two counts: firstly, to shift from using touchscreens/knobs to voice control; then secondly, to shift from using postcodes and addresses to using randomized three-word strings.

Regarding the former, it alters how geodata is transformed into navigational data. Rather than street names, whole addresses, or postcodes being selected from within an addressing database[6] by the user via a touchscreen or dials, three-word strings are spoken and "matched" to the What3words database by LINGUATRONIC. Thus, entirely *new* kinds of vocal data are created – and captured – in every command issued, and every destination uttered. Geodata is literally called into being at the beginning of a navigational task, undergoing a transformation into navigational data as each desired destination or datapoint is enrolled by the "turn-taking machine."

Regarding the latter, it constitutes, what Agre (1995, 186) refers to as "semantic colonization," in which new semantic terms replace others. Not only is one kind of geodata (postcodes) bypassed, but another kind altogether (three-word strings) created, even acting parasitically on the other to "convert" postcodes into What3word locations. This re-datafication process generates an altogether more

5 Thanks to Aikaterini Mniestri for originally alerting me to this, and especially to the colonial practice of place renaming.

6 Such as the UK Royal Mail's Postcode Address File (PAF).

problematic form of value rendered not from local road names or through a standardized (and nationalized) postcode system, but from randomized three-word strings. Whilst it is a strategy that purports to offer representational transparency, it only succeeds in adding further layers, as new navigational "procedures" are designed (Thielmann 2019; Garfinkel 1996).

The Widescreen Cockpit as Customizable Control

The 2018 A-Class not only integrates both voice control and a unique address system into the navigational experience, but also showcases a new dashboard design referred to as a "widescreen cockpit" (Mercedes-Benz 2018c). In this new dashboard, previously separate screens for the instrument cluster and the infotainment system (MBUX) are combined into a single entity operated either via touch or voice. Whilst previously the instrument cluster housing might have been contained behind the steering wheel, it now extends towards the center of the dashboard. Likewise, whilst the infotainment system might have previously existed as either an embedded screen in the center console, or an additional screen attached to the top of the dash, now it stretches across to the driver. In 2021 Mercedes-Benz will launch the "Hyperscreen," a 56inch display stretching the full width of the interior (Hawkins 2021), enabling even deeper forms of customizable control.

In another promotional video, we see a British car enthusiast and YouTuber explain the features of the widescreen cockpit (Mercedes-Benz 2018c). As it begins, he guides viewers through the "three screen set-ups available" to the driver (Mercedes-Benz 2018c, n.p.): two 7-inch displays, one 7-inch display and one 10.25-inch display, and two 10.25-inch displays, meaning "drivers can now customize their display screens." (Mercedes-Benz 2018c, n.p.). In this, there is a double spatial convergence. Firstly, the integration of both instrument cluster and infotainment system into a single entity: the widescreen cockpit. But secondly, a variable spatial convergence (the "three screen set-up") in which the new systems embedded within the interface can be resized. Either users can select two smaller screens (less intrusive), enlarge the media screen (for scrolling through songs etc.), or choose the full, "widescreen" experience.

Figure 3: The new MBUX "Hyperscreen" with three visible screen areas – even wider than the widescreen cockpit

Source: Mercedes-Benz

Next, he describes how drivers can perform this customization: "For example, if you wanted to rearrange the order of these apps [on the infotainment system], you just press and hold on one of the apps, in this case the navigation app, slide it across to where you want it, and then tick confirm and it locks it nicely in place." (Mercedes-Benz 2018c, n.p.)

Here, the media display works akin to the homescreen of a mobile device, in which app icons are presented in a grid-like fashion, movable at the user's discretion. But as he continues, he also highlights the possibility of customizing the instrument cluster, opting for an "understated" style option (Mercedes-Benz 2018c, n.p.). In the 2021 version, the MBUX system will include a "zero layers" feature in which apps will appear "in a situational and contextual way" meaning drivers will not "have to scroll through submenus or [even] give voice commands" (Daimler 2021, n.p.).

Whilst the claim of a "zero layer" display is certainly another example of the strategy of representational transparency, I argue it represents another tendency in vehicle dashboard design. Whilst media functions within cars have been "appified" (Morris and Murray 2018) for a while, instrument clusters and driving-related features have remained off-limits. But as the above, and another later video suggest, "there are many ways to customize your digital dashboard" (Mercedes-Benz 2019, n.p.) that extend beyond media screens and the presentation of apps, and into the representation of vehicle states. That is, to how vehicle data is internally surfaced and transformed into driving data.

This, I argue, constitutes a novel application of Phil Agre's "empowerment and measurement regime" (Agre 1995, 176), in which drivers are "empowered" to customize their digital dashboards, "personalizing" their own driving experience. Yet through novel measurement techniques – datafication by another name – vehicle manufacturers can further "manage" the driving experience. In the 2021 update, this regime is made even clearer and increasingly proactive, as app presentation is "supported by artificial intelligence" through a "context-sensitive awareness [that] is constantly optimised by changes in the surroundings and user behaviour" (Daimler 2021, n.p.). Here, the "MBUX Hyperscreen continually gets to know the customer better and delivers a tailored, personalised infotainment and operating offering before the occupant even has to click or scroll anywhere" (2021, n.p.). What display to show when, and for what purpose, is therefore a kind of situated surfacing in which the MBUX system offers a greater level of customizable control without the need for direct user interaction.

Figure 4: The Hyperscreen visualizing new kinds of driving data (electric vehicle energy "boost" and "recuperation") using a novel "clasp"-style display

Source: Mercedes-Benz

With these technologies, spatial convergence (both fixed, and customizable) leads to an operational convergence, in which both the media display (with app content), and the driving display (with driving data) are presented on the same ontological plane. In this, both are made customizable, not only to the driver's interests, tastes, feelings and passions (Sheller 2004), but also to their driving situations (i.e. going on a family holiday, or driving home from work). Indeed, with the new version of the MBUX, the system will even learn to automatically recommend the vehicle's massage function in cold weather (Daimler 2021, n.p.). Although, as

Mattern (2015) explains, instrument clusters have always enabled additional functionality, for a price, this has typically only concerned *what* can be displayed, not necessarily *how* or indeed *when*. With MBUX we get all three: interchangeable displays that not only show current speed, but also electric motor "recuperation," refueling limits, or the distance to a desired destination; different styles or "skins" for the displays themselves, and context-specificity enabled by artificial intelligence according to who the driver is, and what they're likely to be doing.

Conclusion

Vehicle dashboards are being radically re-designed to represent data differently within the car. With this, vehicle manufacturers are moving beyond the representation of typical features such as fuel levels and speed, or simply the digitization of previously mechanical indicators such as fingers and dials. As vehicles are becoming platformized, new data streams are being generated, sometimes derived from entirely novel operational states. New techniques are being employed to represent this data, and new strategies to convince drivers of their utility.

In this chapter I have considered how the re-design of vehicle dashboards has restructured both navigation and driving through two datafication processes. Firstly, in the transformation of geodata into navigational data, and secondly, in the surfacing of vehicle data as driving data. The spatial and operational convergence of navigation, entertainment and driving features within the car has been a critical enabler of these processes. In order to convince drivers of the value of these design changes, two strategies have been deployed by manufacturers. Firstly, "representational transparency" (Agre 1995, 186) in which new vehicle interfaces are sold as fixes to existing systems deemed "annoying," confusing, or complicated to use. Secondly, "customizable control" in which drivers are afforded greater ability to "personalize" vehicle displays. These strategies, I have argued, are specifically enabled by "re-datafication," in which existing forms of data are transformed into others, actively shaping the experience of driving a car.

To evidence the transformation of geodata into navigational data, I discussed the integration of What3words – a unique addressing system – into a Mercedes-Benz navigation system. Constituting a rather complex operational convergence of multiple systems and functions, navigational capabilities are primarily enabled through voice-activation. Here, geodata is activated as a very particular kind of navigational data, posing both novel practical issues for the driver, as well as constituting a kind of "semantic colonization" (Agre 1995, 186) in which placenames (i.e. geodata) are rendered as randomized three-word strings for the purposes of navigation.

To elucidate the transformation of vehicle data into driving data, I turned to Mercedes-Benz's MBUX system. Here there is an obvious spatial convergence as driving, navigational, and entertainment features are fitted into a single "widescreen" dashboard, or "Hyperscreen." But the MBUX system also demonstrates another more subtle convergence, in which the different display screens of the vehicle dashboard can be customized, enabling driving data to reflect a driver's interests, tastes or personality. As an imminent new version of the MBUX system demonstrates, this customizable control is to be further enhanced through artificial intelligence, enabling a "situated surfacing" of context-relevant functions. These customizable forms, far from being additional elements or features, subtly yet substantially reconfigure the experience of driving. Indeed, rather than a curious side effect, they constitute intended design effects: techniques to transform the driver and their ways of driving.

Thus, this chapter has sought to add to the emerging work on data and platforms, considering how automobiles are being subjected to datafication processes. In the process of transforming both geodata and vehicle data into navigational data and driving data, respectively, new driving experiences are emerging. These, I argue, are worthy of ongoing investigation, as they generate unique effects that help to better understand how datafication is shaping the world-at-large.

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

References

Agre, Phil. 1995. "From High Tech to Human Tech: Empowerment, Measurement, and Social Studies of Computing." *Computer Supported Cooperative Work (CSCW)* 3 (1): 167–195.

Alvarez León, Luis F. 2019a. "How Cars Became Mobile Spatial Media: A Geographical Political Economy of On-Board Navigation." *Mobile Media & Communication* 7 (3): 362–279.

Alvarez León, Luis F. 2019b. "Eyes on the Road: Surveillance Logics in the Autonomous Vehicle Economy." *Surveillance & Society* 17 (1/2): 198–204.

Bolter, Jay D. and Richard Grusin. 1998. *Remediation: Understanding New Media*. Cambridge, MA: MIT Press.

Brown, Barry and Eric Laurier. 2012. "The Normal, Natural Troubles of Driving with GPS." *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* May: 1621–1630.

Brown, Barry and Eric Laurier. 2017. "The Trouble with Autopilots: Assisted and Autonomous Driving on the Social Road." *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* May: 416–429.

Chesher, Chris. 2012. "Navigating Sociotechnical Spaces: Comparing Computer Games and Sat Navs as Digital Spatial Media." *Convergence* 18 (3): 315–330.

Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. "Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb." *Cartography and Geographical Information Science* 40 (2): 130–139.

Daimler. 2018a. "New Infotainment System with Innovative Address System: Daimler Invests in What3words." Accessed January 13, 2021. https://media.daimler.com/marsMediaSite/en/instance/ko/New-infotainment-system-with-innovative-address-system-Daimler-invests-in-what3words.xhtml?oid=32742799.

Daimler. 2018b. "Under the Microscope: LINGUATRONIC Voice Control: 'Hey Mercedes'." Accessed January 13, 2021. https://media.daimler.com/marsMediaSite/en/instance/ko/Under-the-microscope-LINGUATRONIC-voice-control-Hey-Mercedes.xhtml?oid=32997867.

Daimler. 2021. "MBUX Hyperscreen: The Big In-Car Cinema." Accessed January 13, 2021. https://www.daimler.com/innovation/digitalisation/connectivity/mbux-hyperscreen.html.

Dijk, Jose van. 2014. "Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology." *Surveillance and Society* 12 (2): 197–208.

Federal Motor Vehicle Safety Standards (FMVSS). 2020. "Standard No. 101; Controls and Displays." *Electronic Code of Federal Regulations (e-CFR)*. Accessed January 13, 2021. https://www.ecfr.gov/cgi-bin/text-idx?SID=7d443eb75ceba033fed91e90f816b574&node=se49.6.571_1101&rgn=div8.

Garfinkel, Harold. 1996. "Notes Comparing Two Analytic Formats of Occasion Maps of Way Finding Journeys: 'Documentary' and 'Essentially Procedural'." *Unpublished manuscript*, February 24, 1996. Los Angeles, CA: University of California.

Gekker, Alex and Sam Hind. 2019. "Infrastructural Surveillance." *New Media & Society* 22 (8): 1414–1436.

Hawkins, Andrew J. 2021. "Mercedes-Benz Unveils its Absolutely Massive 56-inch 'Hyperscreen' Display." *The Verge*. Accessed January 13, 2021. https://www.theverge.com/2021/1/7/22217267/mercedes-benz-hyperscreen-size-specs-eqs-ces-2021.

Helmond, Anne. 2015. "The Platformization of the Web: Making Web Data Platform Ready." *Social Media + Society* 1 (2): 1–11.

Hind, Sam and Alex Gekker. 2014. "Outsmarting Traffic, Together: Driving as Social Navigation." *Exchanges* 1 (2): 165–180.

Hind, Sam and Alex Gekker. 2019. "On Autopilot: Towards a Flat Ontology of Vehicular Navigation." In *Media's Mapping Impulse*, edited by Chris Lukinbeal, Laura Sharp, Elisabeth Sommerlad, and Anton Escher, 141–160. Stuttgart: Franz Steiner Verlag.

Hind, Sam. 2019. "Digital Navigation and the Driving-Machine: Supervision, Calculation, Optimization, and Recognition." *Mobilities* 14 (4): 401–417.

Hind, Sam. 2020. "Between Capture and Addition: The Ontogenesis of Cartographic Calculation." *Political Geography* 78 (April): 1–9.

Intel. 2018. "Benefits of ECU Consolidation." *Intel Automotive*. Accessed January 13, 2021. https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/ecu-consolidation-white-paper.pdf.

Jenkins, Henry. 2006. *Convergence Culture: Where Old and New Media Collide.* New York: New York University Press.

Kitchin, Rob and Martin Dodge. 2007. "Rethinking Maps." *Progress in Human Geography* 31 (3): 331–344.

Lauriault, Tracey P. 2017. "Open Spatial Data." In *Understanding Spatial Media*, edited by Rob Kitchin, Tracey P. Lauriault, and Matthew W. Wilson, 95–109. Los Angeles, CA: Sage.

Leszczynski, Agnieszka. 2017. "Geoprivacy." In *Understanding Spatial Media*, edited by Rob Kitchin, Tracey P. Lauriault, and Matthew W. Wilson, 235–244. Los Angeles, CA: Sage.

Maalsen, Sophia and Jathan Sadowski. 2019. "The Smart Home on FIRE: Amplifying and Accelerating Domestic Surveillance." *Surveillance & Society* 17 (1/2): 118–124.

Macgregor, Molly. 2020a. "How are What3words Addresses Better than Regular Street Addresses?" *What3words*. Accessed January 13, 2021. https://support.what3words.com/en/articles/3577580-how-are-what3words-addresses-better-than-regular-street-addresses.

Macgregor, Molly. 2020b. "How are the Words Assigned?" *What3words*. Accessed January 13, 2021. https://intercom.help/what3words/en/articles/3577589-how-are-the-words-assigned.

Mattern, Shannon. 2015. "Mission Control: A History of the Urban Dashboard." *Places*. Accessed January 13, 2021. https://placesjournal.org/article/mission-control-a-history-of-the-urban-dashboard.

Mejias, Ulises, A. and Nick Couldry. 2019. "Datafication." *Internet Policy Review* 8 (4): 1–10.

Mercedes-Benz. 2018a. "What3words Voice Navigation: MBUX in the A-Class." Accessed January 13, 2021. https://web.archive.org/web/20200813053637/https://www.mercedes-benz.com/en/vehicles/passenger-cars/a-class/what3words-voice-navigation-mbux-in-the-a-class-2018/.

Mercedes-Benz. 2018b. "What3words Voice Navigation – First to Mercedes-Benz." Accessed January 13, 2021. https://www.mercedes-benz.com/en/innovation/connected/what3words-voice-navigation-first-to-mercedes-benz/.

Mercedes-Benz. 2018c. "MrJWW presents the A-Class (Part 1): The Widescreen Cockpit." Accessed January 13, 2021. https://www.mercedes-benz.com/en/vehicles/passenger-cars/a-class/mrjww-presents-the-a-class-2018-part-1-the-widescreen-cockpit/.

Mercedes-Benz. 2019. "Mercedes-Benz A-Class (2019): How to Personalise the Displays." *YouTube*. Accessed January 13, 2021. https://www.youtube.com/watch?v=pdRTv_LtlWM.

Meyers, Gert and Ine van Hoyweghen. 2020. "'Happy Failures': Experimentation with Behaviour-Based Personalisation in Car Insurance." *Big Data & Society* 7 (1): 1–14.

Morris, Jeremy Wade and Sarah Murray. 2018. *Appified: Culture in the Age of Apps*. Ann Arbor, MI: University of Michigan Press.

National Highway Traffic Safety Administration (NHTSA). 2020. "Instrument Panel Software Incorrect/FMVSS 101/111." *NHTSA*. Accessed January 13, 2021. https://static.nhtsa.gov/odi/rcl/2020/RCAK-20V439-3598.pdf.

Nicas, Jack. 2018. "As Google Maps Renames Neighborhoods, Residents Fume." The New York Times. Accessed January 13, 2021. https://www.nytimes.com/2018/08/02/technology/google-maps-neighborhood-names.html.

O'Kane, Sean. 2020. "Honda Recalls 608,000 Vehicles for Faulty Software." The Verge. Accessed January 13, 2021. https://www.theverge.com/2020/8/4/21354314/honda-recall-software-odyssey-pilot-passport-backup-camera-reboot.

Plantin, Jean-Christophe, Carl Lagoze, Paul N. Edwards, and Christian Sandvig. 2016. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook." *New Media & Society* 20 (1): 293–310.

Pridmore, Jason and Anouk Mols. 2020. "Personal Choices and Situated Data: Privacy Negotiations and the Acceptance of Household Intelligent Personal Assistants." *Big Data & Society* 7 (1): 1–13.

Pridmore, Jason, Michael Zimmer, Jessica Vitak, Anouk Mols, Daniel Trottier, Priya C. Kumar, and Yuting Liao. 2019. "Intelligent Personal Assistants and the Intercultural Negotiations of Dataveillance in Platformed Households." *Surveillance & Society* 17 (1/2): 125–131.

Rose-Redwood, Reuben, Derek Alderman, and Maoz Azaryahu. 2010. "Geographies of Toponymic Inscription: New Directions in Critical Place-Name Studies." *Progress in Human Geography* 34 (4): 453–470.

Rose-Redwood, Reuben, Jani Vuolteenaho, Craig Young, and Duncan Light. 2019. "Naming Rights, Place Branding, and the Tumultuous Cultural Landscapes of Neoliberal Urbanism." *Urban Geography* 40 (6): 747–761.

Sadowski, Jathan. 2019. "When Data is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6 (1): 1–12.

Sheller, Mimi. 2004. "Automotive Emotions: Feeling the Car." *Theory, Culture & Society* 21 (4/5): 221–242.

Shelton, Taylor. 2017. "Spatialities of Data: Mapping Social Media 'Beyond the Geotag'." *GeoJournal* 82 (4): 721–734.

Singh, Ranjit, Chris Hesselbein, Jessica Price, and Michael Lynch. 2019. "Getting 'There' From the Ever-Changing 'Here': Following Digital Directions." In *digitalSTS: A Field Guide for Science & Technology Studies*, edited by Janet Vertesi, and David Ribes, 280–299. Princeton, NJ: Princeton University Press.

Sotoudehnia, Maral. 2018. "Toponymic Checksum or Flotsam? Recalculating Dubai's Grid with Makani, 'the Smartest Map in the World'." In *The Political Life of Urban Streetscapes: Naming, Politics, and Place*, edited by Reuben Rose-Redwood, Derek Alderman, and Maoz Azaryahu, 290–308. London: Routledge.

Stoltzfus, Justin. 2017. "Your Car, Your Computer: ECUs and the Controller Area Network." Technopedia. Accessed January 13, 2021. https://www.techopedia.com/your-car-your-computer-ecus-and-the-controller-area-network/2/32218.

Thatcher, Jim. 2014. "Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data." *International Journal of Communication* 8 (1): 1765–1783.

Thielmann, Tristan. 2018. "Early Digital Images: A Praxeology of the Display." In *Image – Action – Space: Situating the Screen in Visual Practice*, edited by Luisa Feiersinger, Kathrin Friedrich, and Moritz Queisner, 41–54. Berlin: De Gruyter.

Thielmann, Tristan. 2019. "Sensormedien: Eine medien- und praxistheoretische Annäherung." *SFB1187 Media of Cooperation Working Paper Series* 9: 1–10.

US Fleet Tracking. 2020. "Fleet Tracking: Reports." Accessed January 13, 2021. https://www.usfleettracking.com/fleet-tracking/.

What3words. 2019. "All You Need to Know About Mercedes-Benz GPS Voice Navigation Using What3words." Accessed January 13, 2021. https://what3words.com/news/automotive/all-you-need-to-know-about-mercedes-benz-gps-voice-navigation-using-what3words/.

Wilken, Rowan and Julian Thomas. 2019. "Maps and the Autonomous Vehicle as a Communication Platform." *International Journal of Communication* 13 (1): 2703–2727.

Wilmott, Clancy. 2016. "Small Moments in Spatial Big Data: Calculability, Authority and Interoperability in Everyday Mobile Mapping." *Big Data & Society* 3 (2): 1–16.

Winning, Ally. 2019. "Number of Automotive ECUs Continues to Rise." *eeNews Automotive*. Accessed January 13, 2021. https://www.eenewsautomotive.com/news/number-automotive-ecus-continues-rise.

Algorithms Curate Data: Four Perspectives on Data-Based Working Conditions, Using the Example of Route and Job Planning

Annelie Pentenrieder

Users of software services rarely come into contact with data in their everyday lives. There are algorithms working in the background to select, sort, classify and evaluate data for users as part of a process that turns data into information presented on a display. This makes algorithms the gatekeepers that mediate between users and relevant data. The following chapter takes up the question of how algorithms curate data for users while at the same time influencing social arrangements at work and in everyday life. For a theoretical study of data practices, one can enquire into the junctures at which data production and processing data by algorithms play a role for users and consequently must be disclosed.

There is a decades-long tradition in the use of algorithmic means to solve problems by calculating information on the basis of selected data. The omnipresence of algorithmic decision-making we can see in today's world of work and everyday life, however, is new. Algorithms can recommend books, select job applicants, sort e-mails, prioritize information in a wide variety of search engines, and help drivers navigate through city traffic. This is how algorithms have a major impact on the practices and decisions of those who use them and rely on their data analysis. Also the data on which algorithms perform their calculations are collected, processed, and prepared for algorithmic evaluation under conditions that are beyond users' view. And yet data processing does not conclude the moment a data record or a data-collection method (e.g., continuous, real-time data tracking) is set. Therefore, agorithms themselves – the calculation logics under which data are transposed into information for users – must be understood as opaque data practices in their own right. With this in mind, the focus of this chapter is on four different forms of empirical access in which drivers in the logistics sector interact with their software of route and job planning. The four situation descriptions modelled here can be used to illustrate the novelty of working conditions that are subject to algorithm- and data-driven arrangements that are multi-faceted and thus elusive:

> These algorithmic systems are not stand alone little boxes, but massive, networked ones with hundreds of hands reaching into them, tweaking and tuning, swapping out parts and experimenting with new arrangements. If we care about the logic of these systems, we need to pay attention to more than the logic and control associated with singular algorithms. We need to examine the logic that guides the hands, picking certain algorithms rather than others, choosing particular representations of data, and translating ideas into code. (Seaver 2014, 10)

Not only are the data themselves opaque for users, but also the logics that algorithms use to process these data into information. It is not just the results that algorithms recommend that intervene in users' social relationships. There are even social effects stemming from the fact that algorithms, and the underlying data are nearly impervious to scrutiny at this point in time and are used in ways that are opaque. In the following examples, the opacities involved in the process that uses algorithmic calculations to convert data records into information will be explicated, based on work practices of drivers in the logistics sector. In that setting, algorithms take different approaches with regard to the visibilities of information about road traffic and working conditions. This will shed light on the algorithmic architectures in which algorithms structure the social fabric of the people who use them. Using this spatial-theoretical metaphor, I explore the question of how "algorithmically framed perspective" can be "curated" anew, based on the perspective of the user.

Route Planners Invisibly Govern Spatial Access with their Selection of Routes

Anyone who asks a digital route planner to chart a path from Gendarmenmarkt in Berlin to the main rail station will receive a detailed description presenting step-by-step directions indicating where to turn right or left and showing how long it takes to reach the destination. As there are countless ways to get from Gendarmenmarkt to the main station, a navigation algorithm such as the one used by Google selects only a small number of the possible paths. The app developers decide which route will be pre-selected: Should a route be particularly short or quick? Or should a route instead aim to include a reliable arrival time by avoiding roads in advance that often involve unexpected delays due to traffic jams? Even the tiniest of these decisions entails different programming. Even if software providers offer their users different routes to choose from, this selection of options still conceals the technical compromises that the calculation requires. It also covers the developers' own preferences they include into programming, and the data records used to calculate the route. When displayed, a route seems to be the clearly optimal choice from among possible routes, but that is not what it is. The road intersections where a route planner is used are subject to an entire range of assumptions, priorities and conclusions previously made by developers working in a software firm.

As has been argued in Science and Technology Studies, the production of structures and facts that appear to be clear is not a "logical-rational process" or an "objective representation" (Bauer, Heinemann, and Lemke 2017, 13) of a particular subject. Instead, their production must be understood "as a social practice" (2017, 13). According to Susan Leigh Star, technical infrastructures in particular push social development processes to the background – and, with them, the question of who took part in their construction and with which interests (Star [1999] 2017). Classifications, standards, and categorizations that seem clear consolidate power relationships between those who design certain infrastructures and those who use them. This is because the production of infrastructures is inevitably tied to social considerations: "Each category valorizes some point of view and silences another" (Bowker and Star 2000, 5).

For a path between two locations to be recommendable at all on the basis of algorithms –to be calculable, in other words – a wide variety of expectations of a suitable route are simplified on the basis of economic and mathematical criteria. Everyday route planners are based on models in which the cost calculation forms the basis for every concept of space. Space is based on a "mathematizable calculus," and the design of spatiotemporal dimensions becomes a management task and an optimization potential for which mathematical methods and economic ordering structures are used (Neubert and Schabacher 2012, 164–165). When a route is selected, economic factors such as savings of time, money and fuel are prioritized over such harder-to-quantify aspects as convenience, aesthetics and the feelings of the traveller.[1] But even in the case of route properties for which purportedly unique data exists, as in the case of quick or short routes, route decisions cannot be made unequivocally: This has been reported to me by taxi drivers, long-distance lorry drivers and other deliverers whom I interviewed or accompanied in the course of my dissertation project between 2015 and 2019 (Pentenrieder 2020). They report complicated considerations as to which route is fast, short, or economical. Nearly all drivers surveyed used route planners for their journeys, but they always checked the calculations against their own local knowledge and additional sources of information. When choosing their route, they also took into account their own experience – how susceptible a route may be to congestion, for example – or they avoided routes suggested by the route planner that involved bridge underpasses that offered insufficient clearance for an articulated lorry. Or if the "sat nav" device recommended a short distance crossing an avenue, where tree roots

1 On the construction of road data, see: Pentenrieder (2020, 196–205), chapter 6.3: Zum Wert einer Straße: Produktionsketten unveränderlicher Elemente.

typically lift the asphalt to create a "bumpy road," they had to weigh the time savings against wear and tear to the vehicle (field notes, winter 2015). Calculations by a route planner are of little use to drivers in situations like these. Or take another example: During drives lasting several days, a relevant factor for selecting routes for long-distance lorry drivers involved determining which motorways offered places in which to sleep free of charge. This depends on whether roads and rest areas are state-owned or in private hands. Again, none of the information provided by the navigation equipment was of any use, as developers working in software firms know nothing about needs of this type.

In his essay entitled "Information Mythology and Infrastructure," Geoffrey Bowker argues that we should eliminate the "cybernetic narrative" that holds that everything is information, and that realities can be fully represented in models (Bowker 1994, 245). Nor do map data, algorithms and parameters in route planners reflect complete information and universally fitting combinations for suitable routes. They always follow the assumptions about feasibility and interests that can be found in software development firms. Considering long-distance lorry drivers' preferences in rest stops, we can see that the question of which data are available to the algorithm is a function of social and political context. Those who collect data and make them available for algorithmic evaluations take decisions as to which properties *count* and which do not. The inclusion of users' situational knowledge is thus relevant to the effort to adapt algorithmic recommendations to suit the individual situation. Up until now and beyond then, an informed route decision is based on calculable, technically inducible information as well as on the informal knowledge of the users.

And yet informal knowledge bases that users contribute to in the successful application of algorithmic recommendation systems – in this case, the considerations of the drivers themselves – often go overlooked. As "invisible work" (Star and Strauss 1999), they remain hidden when users negotiate the situations of everyday life. Success is instead attributed retrospectively to the provision of algorithmic assistance. The fact that the contribution of an informed user goes unseen owes not least to the steadily growing interpretative power of algorithmic recommendations in comparison with the knowledge of a driver: A taxi driver tells me that passengers have already asked him not to follow his own knowledge of the surroundings but to drive according to the directions shown on the route planner instead (field notes, autumn 2015). Even when I accompany a food supplier in the delivery of his wares, I witness the superior interpretative power of algorithmically calculated data in his everyday work: The software transmits his arrival time to the customer even before the driver, still at the restaurant, has managed to determine the route to take to deliver the food to the customer (accompaniment of a food delivery driver in autumn 2016). Algorithmic calculations forecast and control without offering any way for the stakeholders involved to know the underlying calculation and base data. It is the opacity that makes it difficult for users to review or contradict algorithmic benchmarks in light of other criteria and knowledge levels: "The claim that algorithms will classify more "objectively" (thus solving previous inadequacies or injustices in classification) cannot simply be taken at face value given the degree of human judgment still involved in designing the algorithms, choices which become built-in. This human work includes defining features, pre-classifying training data, and adjusting thresholds and parameters" (Burrell 2016, 3).

Economic and Political Contexts Remain Invisible

A second case in which an algorithmic pre-selection occurs without users' knowledge of the information withheld from them when selecting certain data situations, is provided by the critical urban researcher Ulf Treger, taking Uber, the ride-hailing service, as his example. His work demonstrates the significance of the private development contexts today's software is chiefly programmed. For business purposes, Uber addresses different target groups with specifically tailored map views to influence who receives which spatial access. With internal map views bearing labels such as "Heaven" or "God's view" Uber managers are able to monitor their own business processes and retrieve different layers of information about the urban space (Treger 2018, 241). Because the map information presented to drivers and passengers must necessarily be selected for a limited view in the smartphone app, this selection of information can also be used for business purposes: As Treger points out, the "Hell" map view addresses a group of Uber drivers "who also work for competing companies such as Lyft, the biggest competitor, so that their behaviour can be monitored and controlled: Such drivers are more likely to be offered rides than others in an effort to keep them on duty." (2018, 241–242) Through selection, Uber tailors the "truth content" (2018, 241–242) to certain perspectives on urban space, as is made even clearer in the "Greyball" map view: This map view conceals a secret software program designed to mislead regulatory authorities (Isaac 2017). Between 2014 and 2015, this map view was displayed to people whom Uber had identified as employees of a supervisory authority. Reserving an Uber taxi was made more difficult for this group of people by hiding available Uber vehicles in their immediate vicinity and making them so-called "ghost cars" in order to render it more difficult for the authorities to wield critical control over the company. As journalist Mike Isaac describes in an interview with CNBC, this map view is subject to a special data practice:

> If individuals working for a regulatory authority are "greyballed," they are not banned by the Uber service, as otherwise the app would no longer 'learn' about these people's behaviour: Information about how often the app is opened, and

> about the devices these individuals (e.g., police officers) use to open it, helps reveal the supervisory authorities' tactics. (Isaac 2017)

Along with the first example of the selections made by a route planner, the selection of spatial information shown in an app's map view presents a second possibility of influence in which data and algorithms help determine the social fabric precisely by virtue of the things one does not see in them: by virtue of their opacities. As the map views used at Uber make clear, the focus of a map usage is no longer to capture larger relationships, as noted by Pablo Abend for smartphones in general: "Instead, the viewer is presented with an isolated excerpt that seems to be cut off from any reference." (Abend 2013 118). It is precisely the boundedness of displays that makes it necessary to curate spatial information for users – whether by selecting a route or by selecting a map view. The asymmetrical distributions of information in the "Heaven," "Hell" and "Greyball" views show that technically necessary curation can be accompanied by the opportunity to establish hierarchies in software-structured spatial access. Software already prepares spatial information in different ways for different groups of people: In the process, some groups learn more about a city than others (Treger 2016).

As human geographer Stephen Graham points out, it is particularly software that links places, provides access, and draws boundaries that inscribes new power and knowledge structures as well as inequalities in cities (Graham 2005, 575). It makes a difference which stakeholders deploy those means of digital spatial access, and for which purposes: At the moment, technologies which provide spatial access are strongly intertwined with economic interests, as transnational companies view the urban space as a market in which to sell their technologies (Bauriedl and Strüver 2018, 21). Therefore, any analysis of software concerning spatial constellations, needs a strong focus on social, economic and cultural conditions. In software studies in particular, then, not only is there investigation into what "software is," but also into what "software does" (Mackenzie 2006, 3) and into the material and discursive conditions in which software is embedded today.

Algorithms Assign Tasks Without the Need of Explanations

In the working environment of drivers in the logistics sector, route planners and digital maps are often part of software that additionally coordinates trips, distributes tasks, and controls execution. This is known as "dispatch software." This adds further dimensions on the basis of which algorithms organize visibilities and invisibilities. In a food delivery service as part of the gig economy, for instance, digital platforms or smartphone applications are used to communicate individual orders to the drivers. Essential organizational tasks are outsourced to algorithms, while the operational activities are processed as "repetitive micro tasks" by self-employed workers. (BMAS 2016, 7; Webster 2016; Irani 2015) One social problem with these employment relationships is that the contractors have no guarantees of "adequate economic, social and legal protection" (BMAS 2016, 8).

Min Kyung Lee and her colleagues describe employment relationships such as these, in which gig workers primarily interact with algorithms as superior entities, as "algorithmic management" (Lee et al. 2015, 1603). While the members at the management level remain anonymous, the driver interacts with a fabric of different algorithms. Lee et al. distinguish three dimensions of tasks that algorithms take over in this context: The assignment and issuance of instructions for orders and tasks to the gig worker, the optimization of their work processes associated with the structuring of work steps, and the evaluation of completed tasks (Lee et al. 2015).[2] According to this structure, food delivery staff are assigned their delivery orders using an "instruction algorithm"; then, the second step is to optimize their route to the customer through an interface to Google Maps: with a single click in the same app, drivers can output routes to restaurants and customers in the form of step-by-step instructions. As the third and final step, the app continuously records their speed and location data for evaluation purposes. These data are factored into next order assignment as individual key figures. In addition, target values for punctual delivery are adjusted based on continuous measurements of average speed, i.e. data constantly mediate between drivers' practices and the algorithmically modulated framework in which they operate. These feedback processes are an important element of algorithmic regulation, as they are self-optimizing (cf. Eyert 2020). The algorithm emerges here in the form of a dynamic data evaluation, as it constantly modifies itself based on active driver practices. Unlike gig workers who perform micro tasks online, drivers in the logistics gig economy are a visible part of the cityscape as they perform their algorithm-mediated assignments. In this, they draw attention to problems that affect gig workers in virtual (work) environments as well. Here is how media scientist Carina Lopes describes algorithmic employment relationships in the logistics sector, based on the example of parcel carriers in Spain:

> Algorithms coordinate and indicate then how relations unfold and evolve – what tasks are done, the workflow steps that have been followed and which interactions take place between different parts of the system. Within the intensively computational complex urban systems, everything seems to necessarily start and finish with an algorithm. The parcel about to be delivered enters a spatial field that

---

2 Karen Yeung offers a similar taxonomy for the algorithmic regulation of social arrangements in working conditions. For this purpose, she distinguishes among the dimensions of setting targets, collecting data and modifying behaviours as functions of algorithmic regulation. (see Eyert 2020,1 with reference to Yeung 2018)

> has been calculated and optimised numerous times, all in name of the seamless and efficient workflow towards delivery. It enters a field of action where the bureaucratic algorithm – counting, recording, ordering – meets the automation algorithm – tracking, tagging, deleting, isolating – giving rise to contexts of action that can rapidly evolve and be adapted to environmental aspects such as weather conditions or market demand and supply dynamics. (Lopes 2016, 216)

Algorithms are used to "match" drivers with suitable tasks, as if they were merely one variable among many (2016, 213). In the gig economy, work is structured around the existence of individual orders, as media theorist Felix Stalder points out for digital platforms in general: they create "access to an action space" that "offers opportunities that cannot be found anywhere else" (Stalder 2016, 161). The orders offered on a platform are not attached to a "normative must" but only an "optional can" that either party can retract at any time (2016, 161). Although drivers are free to decide between accepting and declining each incoming order, they have only a limited view of all the factors associated with the respective orders. In the case of Uber, for example, a driver has 15 seconds in which to accept a trip request without seeing where he or she must go – and whether the trip may prove unprofitable (Rosenblat and Stark 2016, 3762).

Opacities like these distinguish algorithmic instructions from instruction made by human superiors. In a bike courier center, I observed the following situation while accompanying the work of a dispatcher:

> During the afternoon of my visit, there were five drivers carrying out incoming delivery orders. As soon as a courier had delivered an order, he or she would use "the radio" to report availability back to the dispatcher. The dispatcher then assigned the driver a spot on the waiting list, with a number between 1 and 5. This assigning of trips continued throughout the afternoon, with one driver repeatedly receiving particularly short delivery orders. (The shorter the delivery route, the less lucrative the order the couriers, as self-employed individuals, can post.) Suspecting that this courier might be 'annoyed', the dispatcher asked all the other couriers if he could assign to this courier an unscheduled, longer delivery run that had just been received. The colleagues agreed, and the driver was given the lucrative order. (Field notes, autumn 2016)

This non-algorithmized instruction situation demonstrated the need for negotiation and argumentation for each standardized assignment procedure with regard to what can be viewed as a fair allocation of work, as well as any situational adjustments that are possible. Because it is presented as indisputable, an algorithmic assignments seems more objectivity – in contrast to biased, non-impartial or distorted human decision-making. But algorithms in the same way assign their tasks in accordance to logics that are subject to negotiation. Unlike human superiors, however, they cannot be interrogated about the rules that govern these assignments. Behind the display, there are technical thresholds, parameters and data inputs that decide how a task is assigned. The algorithmic instructions require not "being in the dispatcher's good graces," as a former bicycle courier described it, but rather a breakdown of mathematical arguments. But these mathematical arguments lie beneath a series of opaque layers that can have technical and organizational, but also political, cultural or economic reasons.

As Louise Amoore has contended, technologies such as "algorithmic modelling" have become the decisive, authoritarian knowledge structures of our time (Amoore 2013, 9). What is problematic about this is that these new objects of knowledge say nothing about the logics and interests inevitably inscribed in them. This results in fractures between different knowledge bases (2013, 9). Programmed by technical development teams, often algorithms and data layers cannot even be explained by direct supervisors; consequently, a delivery driver who interacts with an "instruction algorithm" during trips has no clue of the key figures under which he or she is evaluated or compared with colleagues. If these key figures are output by a "statistical algorithm" as "features" in the context of machine learning, the selection criteria are not even known to the developers, although they may be aware of the spectrum of possible criteria. However, the algorithm "decides" situationally which criteria it deems relevant to the concrete selection, operating on the basis of historical and constantly evolving data. This makes traceability, negotiation, and potential objection of algorithmic instruction very difficult.

Order instructions issued by means of algorithms thus represent a third version of the ways in which algorithms and datasets – which are often dynamically generated and updated[3] – influence the social relationships of their users by virtue of their very opacities. The software determines who receives which assignment and when. It also determines which perspective, and which visibility and invisibility the users will receive onto larger contexts. This algorithmic formation creates "information asymmetries" between drivers and platform providers (Rosenblat and Stark 2016, 3777). But conflicts of interest are systematically resolved in favor of the company that specifies the design of the technology.

3 The ways in which one's own data impact the next delivery order are illustrated by one driver's tactic of playing with his speed to generate more lucrative delivery runs. See Pentenrieder 2020, 131–144, Chapter 5.2.: Wie munitioniert man eine "Weapon of Choice?" [How does one arm a "weapon of choice"?].

Opacities Due to Algorithmic Reshaping

It is not just when job instructions are issued that algorithmic opacities have an effect. Algorithms and the opacities associated with them also influence how workflows are regulated and structured,[4] as I outline in a fourth and final empirical field observation. In autumn 2016, I accompanied a food delivery man on his lunch shift. When picking up a pizza in the restaurant, he only learned the customer's delivery address when the pizza was already hot and in front of him on the counter. Only when he actually set out on his actual route was he able to "unlock the route" and begin planning his path to the customer. The current version of the app does not provide a way to use the time he spends waiting to plan the next route. (Field notes, autumn 2016)

Regardless of whether the app has since been programmed differently, this situation illustrates how the fractures between the programmed arrangements and the knowledge bases of the user can actually hamper the efficiency maxims of logistical arrangements. While algorithmically mediated data can empower new work routines in some cases, they disempower employees in instances in which the algorithms technically reshape employment relationships to such an extent that they structurally reduce and outsource entire areas of responsibility. This prevents not only critical employee input but also constructive employee influences, offered for the sake of the work routine.[5] Here, drivers are permitted to do nothing more than 'fill in the gaps in the automation' (Wischmann and Hartmann 2018, 2). The calling-up of individual knowledge that transcends automation, as Lopes also points out, is growing increasingly difficult, in the interest of "operational efficiency and optimization" (Lopes 2016, 214–215). Work processes allow less and less room for the local knowledge of parcel couriers, for example, who know when certain people are likely to be at home (2016, 214–215). Ethnographers at the University of Hamburg demonstrate that software programming is often not sufficiently geared to the things that would be helpful in a work routine. They refer to the "requirement problem" in computer science, which holds that it is easier to make a software program operational than it is to identify the software solution that a particular situation actually requires. (Brugger et al. 2011, 182) On behalf of "good enough software" (Bialski 2018), digital working architectures simplify socially complex structures in favor of technically feasible programming.

4 In this connection, see the concept of "algorithmic regulation" by Yeung (2018), which was further developed by Eyert et al. (2020) into a framework model.

5 Eva-Maria Nyckel (formerly Raffetseder) and her colleagues reach a similar conclusion with regard to other software interactions (Raffetseder, Schaupp and Staab, "Kybernetik und Kontrolle," 2017, 232).

For decades, feminist technology researchers such as Susan Leigh Star and Lucy Suchman have advocated a "pragmatic sociology of technology" that introduces technology not simply in terms of technical feasibility but also goes on to examine the extent to which it benefits social structures (Star 2017, 35). In lieu of a Turing test that determines whether a machine is perceived as intelligent, Star proposes a Durkheim test that determines whether a machine is perceived as social (2017, 35). Taking into account the rigidities and limitations that technologies and processes of automation inevitably involve, an effort could be made to promote programming that is more oriented to everyday practice and must therefore "reckon with the untidiness of sociotechnical work": "Systems should be tested for their ability to respond to community objectives. [...] A process is deemed to be commensurate if divergent views are factored into the decision-making process in a fair and flexible way" (2017, 131).

Addressing Algorithmic Opacities with User-Centric Perspectives

In different ways, the four examples presented, ranging from the handling of software to route and job planning, demonstrate how software governs which information its users can and cannot see about road traffic and work processes. First of all, route planners grant certain streets the status of making a journey faster than others, yet they do so without disclosing their underlying data and calculation logics. Secondly, map views such as those of Uber restrict the view of the city, prescribing selective spatial access without having to disclose the underlying political and economic considerations involved. There are still more opacities in the context of logistics work: Thirdly, instruction algorithms make it difficult for their employees to critically question standardized job assignments, as even direct supervisors are often uninformed of the technical workings of an algorithm. Fourthly, even constructive contributions from employees are made more difficult in work processes if the task structure is organized in such small steps that one's own knowledge – such as an independent sense of orientation or knowledge of delivery times suitable to the customer – cannot be taken into consideration.

To reduce the instructions shown on the display, information must always be selected for the users – and algorithms evaluate certain datasets for this purpose. By selecting certain information and omitting other information, algorithmic recommendations are always deliberately designed, not objective or neutral. When performing software programming, developers make social decisions with regard to the criteria, user profiles and data that can be used to derive the fit of an individual route or follow-up order. This gives decisions of technical nature a certain social relevance. But what the display conceals are the (social) compromises made on behalf of technically feasible programming, according to which premises, for example, data selection was shaped by means of algorithms.

Precisely because they are unquestionable, algorithms have become new authorities, especially when operating data. They determine which broader relationships users can grasp, understand, and question, and which knowledge of their own they can contribute towards a practice determined by algorithms. Together with all the additional technical conditions that surround algorithms – such as data formats, parameters, thresholds or the limited display space on smartphones – algorithms arrange the visual fields of their users (Pentenrieder 2020, 121–221). Users see the same technical functional logics and organizational backgrounds with which they are interacting only through a "tiny keyhole" (Suchman 2007, 11).

These viewing relationships need to be questioned, because algorithms 'curate' whether, what, when and how their users "catch sight" of a particular piece of information. Algorithmic opacities arise not from a lack of technical knowledge or competences on the part of the users. They arise through the *structural* condition of today's software, which offers too few perspectives onto algorithmic functionalities and data bases for their users.

Spatial concepts can be enlisted to vividly "visualize" these algorithmically defined perspectives in everyday practices. Taina Bucher made this impressively clear in the field of software studies, based on the example of the Facebook News Feed. She shows that, through architectural formations and arrangements, software influences social practices and brings regulatory forces to bear on them (Bucher 2012). In this connection, she projects Michel Foucault's panoptical architectures onto algorithmic arrangements: "An architectural perspective usefully highlights the ways in which spaces are 'designed to make things seeable, and seeable in a specific' way" (2012, 27). Specifically, Bucher enlists John Rajchman's interpretation of Foucault's architecture in the regulation of visibilities: "Architecture helps 'visualize' power in other ways than simply manifesting it. It is not simply a matter of what a building shows 'symbolically' or 'semiotically', but also of what it makes visible about us and within us." (Rajchman 1988, 103). Software installs visual barriers and fields of view in the same manner as spatial architecture installs windows and frames. This is particularly noticeable in interactions with route planners and tasks in the logistics sector, as these practices are determined algorithmically and unfold in spatial structures at the same time. From the point of view of a user at a street corner, one wonders how the software arrives at a route or a task recommendation. This seemingly simple question makes it possible to grow arbitrarily complex where software components and the work steps of the developers involved in the result are concerned. But this initial question makes a major contribution to the discussion of technical conditions: It focuses on the previously limited view of the software user – the view through the "keyhole" and onto the algorithmic result and the datasets – and, from the user's point of view, considers what he or she can "see," and thus know or not know, of the underlying logics of the software. The empirical representation of user interactions prioritizes users' visibilities (in the form of possibilities of knowledge) vis-à-vis the software. In a second step, once users' questions and problems have been elaborated, developers, data scientists and other designers of software can be queried about the technical principles (data, algorithms, parameters, cost functions) underlying the identified opacities that confront users (see methodology in Pentenrieder 2020). Under this approach – of first developing the user perspective and secondly analyzing the technical logics – users can determine what technical information they need to assess algorithmic results and can set their sights on the technical logics essential to this information. What makes this approach decisive is that it *curates* anew what is known about algorithms, this time working from a user-centric perspective.

Analyzing software – not only on the basis of a user-centric construct but also with creative questions in mind about visual relationships – makes a significant difference for a praxeological approach to software analysis. In contemporary urban research, an example of the value of such a shift in perspective towards the object created can be seen in the concept of "human-scale architecture" advanced by the Danish city planner Jan Gehl. In a concept that runs counter to the car-friendly city, Gehl takes the eye level of an individual, completes it with his or her interests, feelings and desires for their own city as habitat and uses this pedestrians' point of view for the planning process. As a result, the functionality and perfection of architecture and solitary buildings are no longer at the heart of urban planning (Gehl 2011). Instead, the planning strategy is defined from the perspective and experience of a pedestrian whose fields of view of his or her city are defined by an eye level at 1.60 m and a walking speed of 5 km/h. As urban research shows, incorporating visual relationships such as these, which pedestrians introduce to their use of urban spaces, significantly alters the planning process and results in cities different to those that have gone before them – "at human scale" (Wang, Sadik-Khan, and Gehl 2012).

To curate informative fields of view for users of a software program, the approach to designing the user experience (UX) calls for a similar shift in perspective – and thus a shift in paradigm. In the 1970s, the development of graphical user interfaces (GUI) marked a radical turning point in the concealment of technical processes.[6] Today, however, one can ask the question in reverse: Which perspectives on technical processes does an emancipatory use of technology now require

6 The development of the "graphical user interface" (GUI) at the research institute Xerox PARC contributed significantly towards establishing personal computers as a mass product on the market in the 1970s. (Chun 2011, 59) It changed the 'dialogue' between user and computer: Users no longer navigated between programs using text-based lines of commands, but now used windows, icons, menus and cursors instead. (Bunz 2019, 76)

for users to be able to develop expertise with regard to algorithmic recommendations – an expertise they can use to check algorithmic calculations, manage and protect personal data and profiles, and control third-party access?[7] A comparison between the user-friendly software design of the 1970s and today's requirements is paradoxical in this respect: While the effort to popularize technology in the 1970s meant that GUI design had to *protect* the user from the technical complexity of software, today's interface design should precisely re-open this protective black box to make informed, responsible or democratic software use possible.[8]

In order to determine which perspectives informed users require, media-theoretical reappraisals of the software itself must be supplemented with the practices of interested users. This is why informed users have also prompted the vignettes formulated above: Their experience shifts the view of the software and the facts of interest that underlie the technology. When describing their everyday lives, long-distance lorry drivers and food couriers also have indirect questions about the production of algorithms and data, along with other basic principles of technology. Some users take this a step further and compensate their limited view by using various reverse-engineering methods and plausibility strategies to reconstruct algorithmic recommendations and relevant software logics (see Chapter 6 in Pentenrieder 2020). These informed and curious drivers try to anticipate the operational principles of software in use, for example by tracking whether it is an algorithm's own logic, movement data the users have generated themselves, or systematic decision-making by a manager or software developer that determines how a task is assigned. Their practices permit conclusions about the social implications that some technical principles involve – principles that thus require social-scientific analysis. This is how emancipatory user practices provide a methodological guide to locating the aspects of a software program that require transparency for users in the first place.[9]

The four scenarios presented here of drivers' interactions with opaque algorithms are based on a combination of spatio-theoretical software analysis and praxeological user research. When users follow a route (1), look at a map on their smartphone (2), accept a delivery order (3) and follow step-by-step work instructions (4), there are algorithms at work that govern the visibilities of data layers for road traffic

7 Cf. the debate around explainable AI: Wachter, Mittelstadt, and C. Russell, "Counterfactual Explanations Without Opening the Black Box," 2017, Kroll, Huey, and Barocas, "Accountable Algorithms," 2017, Spielkamp, Automating Society. Taking Stock of Automated Decision-Making in the EU, 2019.

8 Many thanks to Timo Kaerlein for the inspiring discussion on this topic.

9 In this connection, see methods from participatory design research, such as Suchman 1986/2007.

and working conditions. The empirical case studies manifest the viewing relationship as the source of algorithmic opacities. Particularly the focus on emancipated users reveals architectures in which algorithms govern social structures. Based on user-centric scenarios such as these, specific consideration can be given to how "algorithmically framed fields of view" might be "curated" to make algorithmically conditioned work environments auditable by and accountable to the people who use them.

References

Abend, Pablo. 2013. *Geobrowsing: Google Earth und Co: Nutzungspraktiken einer digitalen Erde.* Bielefeld: Transcript.

Amoore, Louise. 2013. *The Politics of Possibility: Risk and Security beyond Probability*. Durham: Duke University Press.

Bauer, Susanne, Torsten Heinemann, and Thomas Lemke, ed. 2017. *Science and Technology Studies: Klassische Positionen und aktuelle Perspektiven*. Berlin: Suhrkamp.

Bauriedl, Sybille and Anke Strüver, ed. 2018. *Smart City: Kritische Perspektiven auf die Digitalisierung in Städten.* Bielefeld: Tanscript.

Bialski, Paula. 2018. "Hiding in the Stack: Practices of Resistance in Corporate Software Development." Lecture at GfM Annual Conference, September 2018, Siegen.

Bowker, Geoffrey. 1994. "Information Mythology: The World of/as Information." In *Information Acumen: The Understanding and Use of Knowledge in Modern Business*, edited by Lisa Bud-Frierman, 231–247. London/ New York: Routledge.

Bowker, Geoffrey, and Susan Leigh Star. 2000. *Sorting Things out: Classification and Its Consequences: Inside Technology.* Cambridge, MA: MIT Press.

Brugger, Senana Lucia, Katharina Wolter and Steffi Beckhaus. 2011. "Ethnographisch, Praktisch, Gut! Perspektiven für Ethnologen in der Softwareentwicklung am Beispiel eines konkreten Projektes." *Ethnoscripts* 13 (1): 181–198.

Bucher, Taina. 2012. *Programmed Sociality: A Software Studies Perspective on Social Networking Sites*. Oslo: Universität Oslo.

BMAS (Bundesministerium für Arbeit und Soziales) und iit Berlin. 2016. "Foresight-Studie Digitale Arbeitswelt." Berlin.

Bunz, Mercedes. 2019. "The Force of Communication." In *Communication* edited by Paula Bialski, Finn Brunton, and Mercedes Bunz, 51–92. Lüneburg: Meson.

Burrell, Jenna. 2016. "How the Machine 'thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 1–12.

Chun, Wendy. 2011. *Programmed Visions: Software and Memory, Software Studies*. Cambridge, MA: MIT Press.

Dourish, Paul. 2016. "Algorithms and Their Others: Algorithmic Culture in Context." *Big Data & Society* 3 (2): 1–11.

Eyert, Florian, Florian Irgmaier and Lena Ulbricht. 2020. "Extending the framework of algorithmic regulation. The Uber case." In *Regulation & Governance* 1–22.

Gehl, Jan. 2011. *Life between Buildings: Using Public Space*. Washington, DC: Island Press.

Graham, Stephen. 2005. "Software-Sorted Geographies." *Progress in Human Geography* 29 (5): 562–580.

Hartmann, Ernst A. 2015. "Arbeitsgestaltung für Industrie 4.0: Alte Wahrheiten, neue Herausforderungen." In *Zukunft der Arbeit in Industrie 4.0*, edited by Alfons Botthof and, Ernst A. Hartmann, 9–22. Berlin: Springer Vieweg.

Irani, Lilly. 2015. "The Cultural Work of Microwork" *New Media & Society* 17 (5): 720–739.

Isaac, Mike. 2017. "Uber Greyball Programm Evade Authorities." New York Times, March 3. Accessed 06.04.2021. https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html.

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, Laura A. Dabbish. 2015. "Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI'15: 1603–1612.

Lopes, Carina. 2016. *Understanding Relational Locations and Complex Urban Systems: Mapping The Relations Between Computation, Space and Infrastructure*. London: Goldsmiths University London.

Mackenzie, Adrian. 2006. *Cutting Code: Software and Sociality, Digital Formations*. New York: Peter Lang.

Neubert, Christoph, and Gabriele Schabacher. 2012. "Logistik." In *Handbuch der Mediologie: Signaturen des Medialen*, edited by Christina Bartz, Ludwig Jäger, Marcus Krause and Erika Linz, 164–169. München: Wilhelm Fink.

Pentenrieder, Annelie. 2020. *Algorithmen im Alltag: Eine praxistheoretische Studie zum informierten Umgang mit Routenplanern*. Frankfurt am Main: Campus.

Raffetseder, Eva-Maria, Simon Schaupp, and Philipp Staab. 2017. "Kybernetik und Kontrolle: Algorithmische Arbeitssteuerung und betriebliche Herrschaft" *PROKLA* 47 (187): 229–248.

Rajchman, John. 1988. "Foucault's Art of Seeing." *October* 44 (1): 88–117.

Rosenblat, Alex, and Luke Stark. 2016. "Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers." *International Journal of Communication* 10 (1): 3758–3784.

Seaver, Nick. 2014. "Knowing Algorithms." Lecture at *Media in Translation 8*, Cambridge, MA, April 2013.

Stalder, Felix. 2016. *Kultur der Digitalität*. Berlin: Suhrkamp.

Star, Susan Leigh. 2017. "Die Ethnografie von Infrastruktur." In *Grenzobjekte und Medienforschung*, edited by Sebastian Gießmann and Nadine Taha, 419–436. Bielefeld: Transcript.

Star, Susan Leigh and Anselm Strauss. 1999. "Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work." *Computer Supported Cooperative Work* 8 (1–2): 9–30.

Suchman Lucy. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge / New York: Cambridge University Press.

Treger, Ulf. 2016. "Space Making/Space Shaping: How Mapping Creates Space, Shapes Cities and Our View of the World." *Media.CCC*. Accessed December 9, 2018. https://media.ccc.de/v/33c3-7958-space_making_space_shaping.

Treger, Ulf. 2018. "Die Stadt als Bildschirm: Wahrnehmung und Nutzung urbaner Räume durch digitale Kartographie, urbane Dashboards und die Praxis der Navigation." In *Smart City: Kritische Perspektiven auf die Digitalisierung in Städten*, edited by Sybille Bauriedl and Anke Strüver, 237–248. Bielefeld: Transcript.

Wang, Jiangyan, Janet Sadik-Khan, and Jan Gehl. (2012). "The Human Scale.,"Eurovideo Medien.

Webster, Juliet. 2016. "Microworkers of the Gig Economy: Separate and Precarious." *New Labor Forum* 25 (3): 56–64.

Wischmann, Steffen, and Ernst Hartmann (2018). "Zukunft der Arbeit in Industrie 4.0 – Szenarien aus Forschungs- und Entwicklungsprojekten." In *Zukunft der Arbeit. Eine praxisnahe Betrachtung*, edited by Steffen Wischmann, and Ernst Hartmann, 1–7. Berlin: Springer Vieweg.

Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation." *Regulation & Governance* 12 (4): 505–523.

Epilogue

Digitize Again Forever

David Ribes

I am sitting in an office with shelves covered in physical media (see Figure 1). There are papers and folders and binders, CDs and tapes everywhere, stacked on top of each other, mixed and matched. Note the stamps, which are intended to label and categorize paper; and note the papers bursting from binders. Note the specialized drawers for CDs; and note the stacks of CDs on top. Despite the apparent mess this is not a disordered space, at least not for my interlocutor who knows, in broad stroke, which pile is which. She strides through the piles with confidence.

This is the data management center's main office for the Multicenter AIDS Cohort Study (MACS). Since 1987 that data management center has been curating, preserving, cleaning, combining, and provisioning the data of the MACS. They are also responsible for digitizing the MACS's data again:

> We digitized these data in the early 2000s. You can see the CDs behind you [she gestures to shelves behind me], which we have to *digitize again* soon because CDs only have a shelf life of about 10 years. They are overdue. (emphasis added)[1]

I've sat in other rooms like this before. Since the MACS has three other geographic sites across the US, there are also three other rooms that look somewhat akin to this one. The piles, the heterogenous media, the definitive sense of order which is nevertheless inscrutable to my eyes.

This room is in Baltimore, Maryland, USA, but this detail doesn't matter very much to the tales I will tell here. I find that the rooms of information managers across my different research projects all look much the same. Even when it's not the MACS, and even when it's not about AIDS, I've still been in rooms that look much like this.

1 Anonymized interview with MACS Information Manager, October 2011.

Figure 1: Center for Analysis and Management of MACS Data

Source: Photograph by the author, all rights retained.

The MACS was founded in 1984. Then, data were first inscribed on paper forms, quickly transferred to long magnetic tapes and (for them) startlingly compact floppy disks. Those data on paper or disk were shared across its four geographic sites by fax and by mail, but only after first phoning someone to confirm that the data were 'coming by fax' or 'in the mail'. That was all a long time ago and a lot of things have changed since, such as the disease they study (fatal in 1984, a chronic parasite today), their scientific instrumentation (which regardless, almost all end up outputting data), and of course all the media that they use to store these data. Nothing has stayed the same, except of course that the MACS keeps producing more data, and then digitizing it, and then digitizing it again.

After the $5\frac{1}{4}$ inch floppy disks there were the even more compact $3\frac{1}{2}$ inch disks, hard drives (which were also placed in the mail for a time), slightly larger but somehow more compact discs, and of course The Internet. Of course, The Internet did not pass these information managers by, they kept up. But despite its tidy name The Internet is not one thing, and it has done nothing to end the parade of digitizing their data again. There have been shared networked drives; carefully guarded servers run by departments, colleges or the university; even more carefully guarded private provider servers; and of course a recent bevy of Cloud services. All of this has been The Internet, and none of it has slowed the MACS' efforts at digitizing their data again:

> The last time a vendor told me they were going to "digitize our data" [curling her fingers around the words], I guffawed out loud. I know it was rude, I couldn't help it, I'd heard it a few times before…. [chuckles] Anyways, we did end up hiring them, and they did digitize our data again.[2]

How Digital Can We Get?

Here I reject what I call the saltationist model of digitization (James 1909; Latour 1999). My argument is that being digital is itself not a binary state, a singular jump from one condition to another, despite how much the digital itself insists that is the case. Instead, I consider digitization to be hungry, bottomless, always looking for its next meal, forever dissatisfied with everything in the past, including its own past efforts of digitization. The children of the digital, those dubbed 'born digital', are not exempt, they are just as tasty, perhaps more so, and just as subject to digitization's endless maw. Like Saturn, digitization will consume them too, and then once again.

I have thought that digitization is an endless hungry maw for some time now. I have been studying scientists digitizing their data again since about the early 2000s. As we saw above, in the past I've worked with biomedical scientists and their data; I've also worked with geologists, ecologists and physicists digitizing their data again, amongst others.[3] I should know by now that digitization is not a binary state, more than for any other reason because sometimes I have returned a decade later to the very same group of scientists only to find that they are once again digitizing their data.

Often, they don't call it digitizing their data, there are other words. Such as 'implementing' a metadata standard, or 'registering' data to a computational ontology, or 'harmonizing' across heterogeneous data. In this I am not fooled though, I am theoretically sensitized (Glaser 1978) enough to recognize a common pattern when, regardless of what they call it, really they are just digitizing their data again. In this, the central theoretical term of this edited volume, "datafication," too smells to me suspiciously like digitization again.

But even with all this critical distance, I still find myself regularly duped by the promissory pitch: 'just try it once more' digitization coos in my ear, 'this time will certainly be The One'. The promise of digitization is that, going forward, things will never be the same again. For digitization, The One is the occasion on which being digital has been achieved, forever. This is how we often speak of digitization, a singular transformative leap from some anterior condition (such as "ana-

2 Anonymized interview with MACS Information Manager, October 2011.

3 Ribes and Polk 2015; Ribes and Bowker 2009; Ribes et al. 2012.

log," sometimes "material," or perhaps "qualitative") to a new state of digital being. It is a one-way trip to becoming binary and symbolic, letters of light, electricity, and magnetism. Data.

I know this claim cannot be true. I argue against The One transformation in many of my papers, (including this one, such as with the bit of ethnographic evidence above). And yet still, I am not immune to the siren call of digitization. I often head straight towards it, convinced.

A small example via a true personal story: Just recently, about 10 years ago now, I digitized my music collection. My collection was mostly CDs. As these discs appeared to be on their way out the door I adopted a portable MP3 player, and then soon, a smartphone to replace that too. As a practice, digitizing my music collection looked like this: I would insert the CD into the drive and using software ("ripper") converted the CD's native format (WAV) into the more compact, but slightly lossy, MP3. Then, I repeated this task with the next CD. It took weeks.

Obviously, the CDs were "already digital," WAV files are data (Sterne 2012). Even still – and equivalent to the data CDs of the MACS above – the software I used called it digitization and so too did the guides I looked up online for how best to "Digitize your music collection." Technically, something digital cannot be digitized, in practice we digitize again all the time.

Despite all the skepticism I have communicated above, I really did digitize my music thinking that this time was The One. That, after all the musical media transformations I had gone through (…vinyl, tape, CD…), now I would no longer need to transition again because now my music would be digital. Files of light and magnetism that I could take, no matter what, with me to wherever the digital went next. Obviously, I was wrong; but it's *how* I was wrong that I think is of interest. Digitization proved surprisingly discontinuous from its own past.

In my defense, I was not wholly naive. I did bring to bear some of my academic craft to the task. Such as thinking to myself that certainly the future would still involve a great deal of maintenance and repair, a practical labor of transition. Perhaps MP3s would go out of date, and I would have to spend some time converting (i.e., digitizing again) those files to the new format. But I still thought that the files themselves were now eternal, just a matter of converting on to the next thing. In this I was wrong; this is not at all what happened.

That digitized music collection – stored as hand-crafted MP3 files – today sits saved on a hard drive that itself sits on top of the original CD collection, in my garage, in a box, untouched for nearly a decade (see Figure 2). My prediction above – that this digitization was finally The One, or that I would simply convert these files to the next format – did not pan out as such. I did not change these digital files into anything because the next form for digitization turned out to be a completely different kind of digitization, wholly exceeding my predictions as is often the case with the wily ways of digitizing again.

Figure 2: The author's digital CD music collection, digitized again onto a hard drive, which resides in this box, in the garage, untouched for a decade, until he took this picture.

Today my music collection is provided to me through The Cloud. I pay a monthly fee to have access to the music I listen to. I never digitized my collection, at least not in the sense of 'uploading' my MP3s to The Cloud or some other form of direct converted continuity for these files. Instead, in this case, digitizing my music again involved selecting albums and tracks from a vast drop down list of "almost all music." It took weeks.

I'd be comfortable saying that this new collection is equivalent to my old collection. But it has also grown, shrank, and some tracks just don't sound the same. The Cloud, as with CDs and MP3s, and all other digitizations that came before, insists that this is the end-state for my music collection, forever.

The Saltationist Model of Digitization

A saltation is a jump, perhaps a particularly vigorous jump. In the sciences saltation is used metaphorically to indicate a dramatic, singular transformation, such as in biology where a saltation is an abrupt evolutionary change. But while there *may* be saltations in evolution, digitization does not occur as a singular decisive and transformative leap. Instead I consider digitization to be a bottomless discontinuous chain, and I name my oppositional foil, the saltationist model of digitization.

I draw this term from actor-network theorist Bruno Latour, who in turn borrows it from Pragmatist philosopher William James (James 1909; Latour 1999). Latour and James too use "saltation" to denote their foil: a vast set of epistemological and ontological commitments that we might call representationalism or objectivism (Ribes and Sutherland 2021); a position that asserts a world "out there," its representation "here," and the gap between these two as the central matter for concern. Saltation in Latour and James broadly evokes classical modernist philosophical distinctions between substance and form, matter and mind, and most notably world and representation. I will not tackle these vast theoretical commitments in this little essay, but I too intend my use of saltation to evoke how digitization has often been miscast in the idiom of representationalism.

Here, I use saltation more modestly to define the model that casts digitization as an historical binary, a once-this-then-that, single-instance-transformation. Instead I argue that *digitization is, in principle, a bottomless chain of discontinuous equivalencies centrally concerned with continuing to exist.*

I use the term *equivalency* in the sense imparted to that term by actor-network theory: one thing taken to be the same as another. More strongly than 'taken to be', that if something or someone were to dissent (Latour 1987) as to that similarity, the equivalency would fight back. Equivalencies are real, in the Pragmatist sense of that which resists. When digitizing the data of the MACS from magnetic tape, to disk, to disc, to The Internet, it is always the same, equivalent. If MACS scientists did not think equivalency had been achieved, then it would not be digitization at all, merely dangerous and disposable garbage data. If I, or anyone else, was to suggest that those data are not the same, the MACS would fight back. The equivalencies of the before-and-after of digitization are very real, but each era of digitization is discontinuous from the last.

I use the term *discontinuous* in the sense that digitization-next regularly proves to be distinct from digitization-before. Digitization itself denies this, it will tell you that it is always just ones and zeros, which is what it was before and what it will be next. But approached as practice – that is, respecified (Garfinkel 1991) – digitization-next is rarely like digitization-before.

Recall that when I first digitized my music collection (ripping) it turned out to be quite different than the second time (selecting music from The Cloud). So too for scientists and all their legacy data. This next tale is no different in its form, even though it is more technical (by which I mean, less easy than ripping): Once I observed geologists over months as they sought to digitize their data again, on this occasion by "registering" their data to a "computational ontology." At the beginning of this process they already had some of the best organized data out there: in a relational database, with plenty of descriptive metadata, readily available online via a Web service. Even still, something was lacking, such as being able to conduct a semantic search, something ontologies would allow them to do. But registering to an ontology demanded something additional, something more, i.e., formally encoding their geoscientific knowledge into the predicate logics of the Ontology Web Language (OWL), and then registering their data, column by column, to that domain-specific ontology. In the paper I collaboratively wrote from watching these geologists, we argued that geoscientists had to "learn" (Ribes and Bowker 2009) to articulate their informal knowledge in the predicate logic, because even though these geologists were already highly data literate, and their data were highly organized, open access and online, in the face of these computational ontologies they still had to learn how to digitize their data again, a discontinuous practice from what they had done before.

Lastly, I use the term *bottomless* somewhat colloquially, like a Bottomless Mimosa: in principle there is always more to be had, in practice there are concrete limits; which is also how I mean *forever*. Digitization is bottomless in that there will always be another instance of digitization just around the corner, demanding, needed. The next form of digitization will always insist on something new, something more. Data will need to be relational or maybe object-oriented; there will be a need for more or better metadata; they must be uploaded, online, registered, open access, reproducible, differentially privatized… The point is that sooner than later data are 'overdue' and at risk of being lost. Respecified as practice, the next digitization will always be different than what came before, even while digitization will always insist it's just a matter of ones and zeros, a one-off transformation like before again.

Clearly, there is some element of a corporatist motive here – as with the vendors for the MACS who forever promise a final digitization only to return a few years later with a new one; or, as with The Cloud provider that I now pay monthly for my music. There is money to be made by promising The One transcendence even while digitizing again, forever. But there is more to digitization's bottomless hunger than greed; it is also about continuing to exist.

In principle the data of the MACS, stored on CDs, or my music collection, also stored on CDs, could continue on indefinitely, with the sort of maintenance and repair I suggested above: tweaks along the way. But in practice I think these data, left on CDs, would cease to exist, at least practically. Consider, if today an HIV scientist requested data from the MACS and was told that those data would arrive by mail, on a CD, in a comma-separated flat-file, in a couple of weeks; I think that researcher would at least consider looking elsewhere for data, such as via The Internet. Even if 20 years ago they would have regularly waited two weeks for comma-separated data, today many would not. Personally, I have become used to waiting no more than a second for my music, even if I know that the track that plays from The Cloud may not quite be the one that would have played from the CD in my collection. In principle, well-kept data are eternal, in practice they must be digitized again, not because the data have changed, precisely, but because they cannot stay the same while everything else does not.

To say that digitization is a bottomless chain of discontinuous equivalencies which are centrally concerned with continuing to exist, is a long way of saying digitization is a parasite – in the sense imparted to that term by poet-philosopher Michel Serres. In order to continue to exist, and despite all promises of a one-time transcendence, digital things, such as data, but also everything and anything else, must be interrupted, transformed and renewed, now, and then again later, forever:

> a new obscurity accumulates in unexpected locations, spots that had tended toward clarity; we want to dislodge it but can only do so at ever-increasing prices and at the price of a new obscurity, blacker yet, with a deeper, darker shadow. Chase the parasite – he comes galloping back, accompanied, just like the demons of an exorcism, with a thousand like him, but more ferocious, hungrier, all bellowing, roaring, clamoring. (Serres 2013)

The Parasite can be read as an essay on information – "In the beginning was the noise," Serres writes – but it can also just as easily be read as about datafication, The Internet, The Cloud, or digitizing. They are not that different. More accurately The Parasite is about information again, datafication again, or uploading again, which are all equivalent discontinuities for the same thing: digitize again forever, "the chain seems unending" (Serres 2013).

References

Garfinkel, Harold. 1991. "Respecification: Evidence for Locally Produced, Naturally Accountable Phenomena of Order, Logic, Reason, Meaning, Method, etc. in and as of the Essential Haecceity of Immortal Ordinary Society (I) – An Announcement of Studies." In *Ethnomethodology and the Human Sciences*, edited by Graham Button, 10–19. New York: Cambridge University Press.

Glaser, Barney. 1978. *Theoretical sensitivity*. Oakland, CA: University of California Press.

James, William. 1909. *Pragmatism and four essays from the meaning of truth*. Cambridge, MA: World Publishing Company.

Latour, Bruno. 1987. Science in Action: How to Follow Scientists and Engineers Through Society. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1999. *Pandora's Hope : Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Ribes, David, and Geoffrey C. Bowker. 2009. "Between Meaning and Machine: Learning to Represent the Knowledge of Communities." *Information and Organization* 19 (4): 199–217.

Ribes, David, Steven J. Jackson, R. Stuart Geiger, Matthew Burton, and Thomas Finholt. 2012. "Artifacts that Organize: Delegation in the Distributed Organization." *Information and Organization* 23 (1): 1–14.

Ribes, David, and Jessica Beth Polk. 2015. "Organizing for Ontological Change: The Kernel of an AIDS Research Infrastructure." *Social Studies of Science* 45 (2): 214–241. Accessed September 27, 2021. https://doi.org/10.1177%2F0306312714558136.

Ribes, David, and William Sutherland. 2021. "Representationalism and Digital Imagery." In *Postphenomenology and Imaging: How to Read Technology*, edited by Samantha J. Fried and Robert Rosenberger, 283–292. Lanham/Boulder/New York/London: Lexington Books.

Serres, Michel. 2013. *The Parasite*. Minneapolis, MN: University of Minnesota Press.

Sterne, Jonathan. 2012. *MP3: The meaning of a format*. Durham, NC: Duke University Press.

Authors

Marcus Burkhardt is Postdoctoral Researcher in the Media Studies Department at the University of Siegen. Together with Karin Knorr Cetina, he heads the subproject "Agentic Media: Formations of Semi-Autonomy" in the DFG-funded Collaborative Research Center 1187 "Media of Cooperation". He has a background in media studies, philosophy and computer science and received his doctorate at Justus-Liebig-University Gießen with a thesis on the media theory of digital databases. His research interests include the history and theory of digital media, especially the logi(sti)cs of database technologies, big data and algorithmic media as well as digital methods. More info: https://www.marcus-burkhardt.net

Igor Eberhard is a University Lecturer at the Department of Social and Cultural Anthropology, University of Vienna, Austria. His research focuses on skin studies, attribution of criminality and deviance, research ethics, collections, on which he also holds a Ph.D. He is head of the department's ethnographic collection. Currently Eberhard is also project manager of the Ethnographic Data Archive (eda), together with Wolfgang Kraus since 2017 at the University Library Vienna. Website: eda.univie.ac.at.

Emma Garnett works at the intersections of sociology, anthropology, and STS and is a Senior Research Fellow in the Social Science and Urban Public Health Institute at King's College London. Using ethnography and other social science methods, her research employs collaborative, interdisciplinary formats to investigate how airs, atmospheres, and breathing are associated with environmental and human health. Recent projects concern the changing configurations of science, medicine, and public health in response to innovations in air pollution sensing, and the political and organisational impacts of COVID-19 on waste management in the UK.

Daniela van Geenen is a Ph.D. candidate at the DFG-funded Collaborative Research Center 1187 "Media of Cooperation" at the University of Siegen, after receiving her MA from Utrecht University. Her research lies at the interface of media studies and science and technology studies (STS), with particular focus on (critical) data studies, software, device, and infrastructure studies. In her Ph.D. project she investigates the sociomaterial organization, (data) practices, and values in/of urban sensing, paying specific attention to cases of environmental sensing in relation to mobility monitoring and planning. She (co-)authored several journal articles and book chapters and is currently co-editing a special issue in the SAGE journal *Convergence* on "Critical Technical Practice(s) in Digital Research." Daniela is also a lecturer in Data Journalism and Visualization at the University of Applied Sciences Utrecht.

Carolin Gerlitz is Professor for Digital Media and Methods at the University of Siegen and member of the Digital Methods Initiative Amsterdam. Her research explores the various intersections between digital media, methods and data, with a specific interest web economies, platform and software studies, apps, sensory media, quantification, automation, social media and inventive methodologies. She is Co-Leader of the Center for Digital Methodologies in Media, Language and Technology Research and deputy speaker of the DFG-funded Collaborative Research Center 1187 "Media of Cooperation".

Flis Henwood is an Emeritus Professor of Social Informatics in the School of Humanities and Social Science at the University of Brighton, with a background in STS and medical sociology. Her research focuses on understanding the design, development and use of information and communications technologies (ICTs) in everyday work and life settings, with particular emphasis on how such technologies mediate and shape health and social care practices. Examples of her research include the development and use of electronic patient records in a number of settings, as well as the informational practices and support needs of lay people in a variety of contexts. Her recent research focused on the development and evaluation of a mobile phone app for stable HIV patients, and on self-monitoring. More info: https://research.brighton.ac.uk/en/persons/flis-henwood

Sam Hind is Research Associate at the DFG-funded Collaborative Research Center 1187 "Media of Cooperation" at the University of Siegen. He is co-editor of *Time for Mapping: Cartographic Temporalities* (Manchester University Press, 2018) and co-author of *Playful Mapping in the Digital Age* (Institute for Network Cultures, 2016). He completed his Ph.D. on protest mapping in the Centre for Interdisciplinary Methodologies (CIM) at the University of Warwick, UK. His research interests include navigational technologies, sensing and the sensor society, algorithmic decision-making, and automotive cultures. Since 2017, his research has focused on the design, development, and testing of autonomous vehicles.

Araa al Jaramani holds a Ph.D. in semiotic trends and criticism from the University of Damascus. She worked as a researcher at the Arab Reform Initiative in Paris, as a Postdoctoral Researcher at the Centre for Sociology (GERME), Université Libre de Bruxelles, at the Department of Media and Culture Studies, Utrecht University, at the Van Vollenhoven Institute for Law, Governance, and Society, Leiden University, and as a Head Researcher at Save the Children Netherlands on the program Kifak. She founded "Syrische Vrouwen in Nederland (SVNL)". She has published about socio-cultural issues, especially migration and gender.

Timo Kaerlein is Lecturer at the Institute for Media Studies at Ruhr University Bochum. From 2018 to 2021 he was Research Coordinator at the DFG-funded Collaborative Research Center 1187 "Media of Cooperation" at the University of Siegen. His research focuses on the theory, history, and aesthetics of interfaces, infrastructures and practices of urban sensing, as well as theories of embodiment. He holds a Ph.D. in media studies from the University of Paderborn on smartphones as embodied digital technologies and post-cybernetic control infrastructures.

Helena Karasti is an interdisciplinary scholar with longitudinal research interests in ethnographic practice studies, participatory design, computer supported collaborative work and information infrastructure studies. Her main empirical domains of investigation include healthcare and environmental research. She has published in the topic areas of scientific data/information management, research infrastructure formation, infrastructuring, integrations of ethnography and participatory/collaborative design, participation, and researcher reflexivity. Karasti has worked as professor at University of Oulu and Turku (Finland) and Luleå University of Technology (Sweden). She has been guest professor in the Department of People and Technology at Roskilde University (Denmark), visiting scholar in the Institute for Advanced Studies on Science, Technology and Society (IAS-STS) in Graz (Austria), and postdoc in the Department of Communication at University of California, San Diego (USA).

Wolfgang Kraus is Associate Professor at the Department of Social and Cultural Anthropology, University of Vienna, Austria. He has done extensive ethnographic research on tribal identity and organisation as well as oral historical knowledge among Imazighen/Berbers in central Morocco. Among his research interests are ethnography, kinship, visual and multimedia anthropology, data management and research ethics. Currently he is leading the Ethnographic Data Archive (eda) that he established together with Igor Eberhard since 2017 at the University Library Vienna. Website: eda.univie.ac.at.

Danny Lämmerhirt is Research Associate at TU Dortmund in the DFG-funded Collaborative Research Center 1472 "Transformations of the Popular" where he studies how digital health apps change the doctor-patient relationship. Blending media studies with STS and sociology, his research explores the politics of participatory data infrastructures, data governance, quantification, and valuation. His Ph.D. project develops a media studies perspective on 'citizen-centered' health data intermediaries and how they enable cooperation, governance, and the valuation of health data along multiple, possibly conflicting, interests. More info at: www.dannylammerhirt.org.

Gaia Mosconi is research assistant at the University of Siegen. She holds a Master in Sociology and a second Master in "Smart Community Design and Management". Since 2016, she is a member of the CRC1187 "Media of Cooperation" where she works as project manager, conceptual designer and researcher in the INF project. In INF, Gaia investigates the field of RDM with a practice-based approach, studying how collaborative research data practices and infrastructures are changing under the auspices of the Open Science agenda. Finally, she is also actively involved in the design and development of a Research Data Infrastructure 'in the making'.

Claudia Müller is Deputy Professor for "IT for the Ageing Society" at the Institute of Information Systems at University of Siegen and Professor at Kalaidos University of Applied Sciences Switzerland, Department of Health Sciences, Zürich. Claudia's research interests are in inter- and transdisciplinary approaches between applied informatics and ageing and health research. Thematically, she is researching sociotechnical systems in the health and ageing domain, such as care robotics, accessible workplaces, IT in caring communities and hybrid IT systems in healthcare. Methodologically, her research interest lies in participatory and community-based co-design methods. More info: https://italg.wineme.uni-siegen.de/en/

Nathaniel O'Grady is Lecturer in Human Geography and Disaster at the Humanitarian Conflict Response Institute, University of Manchester. His research focuses primarily on the cultural-political implications that arise where new technologies become entangled within practices of governance.

Annelie Pentenrieder is a Scientific Advisor at the Institute for Innovation and Technology (part of VDI/VDE-IT), Berlin. She works on participatory technology design and questions of digital sovereignty in industrial environments. In her dissertation, she analyzed road traffic route planners, and how human actors engage in the critical evaluation of algorithmically-generated information in everyday, working routines.

Volkmar Pipek is Professor for Computer Supported Cooperative Work and Social Media with the Institute for Information Systems at the University of Siegen. Previously, he worked at the Universities of Bonn and Oulu. His research focuses on

(IT-supported) cooperation between humans and machines, as well as appropriation practices of such systems. He currently chairs to the board of trustees of the International Institute for Socio-Informatics (IISI).

Sandra Ponzanesi is Chair and Professor of Media, Gender and Postcolonial Studies at the Department of Media and Culture Studies at Utrecht University. She was PI of the ERC project "Digital Crossings in Europe: Gender, Diaspora and Belonging" (CONNECTINGEUROPE), which investigated the relation between migration and digital technologies, focusing on how digital diasporas are created and performed in relation to issues of gender, ethnicity, and affective belonging. She is director and founder of the PCI (Postcolonial Studies Initiative). She has published widely in the field of media, postcolonial studies, digital migration, and cinema with a particular focus on Postcolonial Europe from a comparative and interdisciplinary perspectives. More info: https://www.uu.nl/staff/SPonzanesi/Profile

Dave Randall is a senior professor at the University of Siegen in Germany and visiting professor at Linnaeus University in Sweden. He is co-author of four monographs including 'Fieldwork For Design' with Richard Harper and Mark Rouncefield, and 'Choice' with Richard Harper and Wes Sharrock, and two edited collections. He has published in excess of 100 peer-reviewed papers and has undertaken consultancy work with organizations such as Vodafone; Hitachi; Orange; Microsoft; a hospital trust; Xerox; and the Children's Society. His primary interest lies in ethnographic practice in the context of CSCW (Computer Supported Cooperative Work) and HCI (Human Computer Interaction).

David Ribes is Associate Professor in the Department of Human Centered Design and Engineering (HCDE) and Director of the Data Ecologies Lab (deLAB) at the University of Washington. He is a sociologist of science and technology who focuses on the development and sustainability of research infrastructures (that is, networked information technologies for the support of interdisciplinary science), their relation to long-term changes in the conduct of science, and transformations in objects of research. His current research investigates the emerging institutions of *data science* at multiple scales, such as changing scientific practices, budding regional or national organizations, and novel public-private partnerships. David is regular contributor to the fields of STS and information studies. His methods are ethnographic, archival-historical and comparative. See davidribes.com or dataecol ogi.es for more.

Minna Ruckenstein works as Professor at the Consumer Society Research Centre, University of Helsinki. She directs The Datafied Life Collaboratory that studies processes of digitalization/datafication by highlighting emotional, social, political and economic aspects of current and emerging data practices. The disciplinary underpinnings of the work range from anthropology of technology, science and technology studies to media and communication. Currently funded research projects focus on re-humanizing automated decision-making, algorithmic culture, and everyday engagements with algorithmic systems in Helsinki and in Shanghai and Hangzhou.

Gerwin van Schie is Lecturer at the Department of Media and Culture Studies at Utrecht University where he is completing the NWO-funded Ph.D. project "Datafication of Race and Ethnicity in the Netherlands: Investigating Practices, Politics and Appropriation of Governmental Open Data". In his research he combines insights from postcolonial studies, critical race studies and critical data studies in the investigation of Dutch governmental data infrastructures that make use of various data relating to race, ethnicity, nationality, and migration background.

Kyle Stine teaches film and media studies at Johns Hopkins University. His writings on cinema and technology have appeared in *Critical Inquiry*, *Discourse*, *Grey Room*, *October*, and the *Journal of Cinema and Media Studies*. He is coeditor with Axel Volmar of *Media Infrastructures and the Politics of Digital Time: Essays on Hardwired Temporalities* (Amsterdam University Press, 2021).

David Struzek (M. Sc.) works as a Research Assistant at the Institute of Information Systems at the University of Siegen in the area of "IT for the aging society". As a Ph.D. student he is interested in the question how social participation through technology can be supported in urban and rural areas by health-promoting technology solutions that are developed with participatory methods. David's further research interests are in the area of human-computer interaction, user experience, usability and accessibility, and creative technologies. https://italg.wineme.uni-siegen.de/team/david-struzek/

Tommaso Venturini is a Researcher at the CNRS Centre for Internet and Society, Associate Professor at the Medi@lab of the University of Geneva, and founding member of the Public Data Lab. He was a Researcher at ENS Lyon, Research Fellow at INRIA, Lecturer in digital humanities at King's College London and Coordinator of the Médialab at SciencesPo Paris. Tommaso's research lies at the intersection of media studies, science and technology studies, and digital social sciences. They focus on digital methods, online debate, attention cycles and the secondary orality of platforms. Find out more: http://www.tommasoventurini.it/

Axel Volmar is a Postdoctoral Researcher at the DFG-funded Collaborative Research Center 1187 "Media of Cooperation" at the University of Siegen. From 2014 to 2016, he was a Mellon Postdoctoral Fellow in the Department of Art History and Communication Studies at McGill University. His research interests include the history, theory, and praxeology of digital media, sound studies, format theory, and digital time. He is co-editor of several books, most recently FormatMatters.Standards,Practices,andPoliticsinMediaCultures (meson press, 2020) and MediaInfrastructuresandthePoliticsofDigitalTime.EssaysonHardwiredTemporalities (Amsterdam University Press, 2021).

Kate Weiner is a Senior Lecturer in the Department of Sociological Studies at the University of Sheffield. Her work sits at the intersection of STS and medical sociology, focussing on developments in the biomedical sciences, and everyday health practices. She is interested in the implicit models of the social world that inform knowledge production, technological developments, and health policy, and how these intersect with practices in clinics and in everyday life. She has researched in the areas of genetics, heart disease and patient's organisations, everyday medication and eating practices, race and prescribing, and most recently on self-monitoring. More info: https://www.sheffield.ac.uk/socstudies/people/academic-staff/kate-weiner

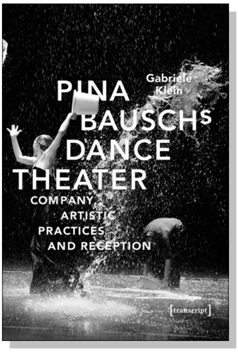
Catherine Will is a Reader in the Sociology of Science and Technology in the department of sociology at the University of Sussex. Her research examines social and health care organisations and technologies, drawing on STS and political sociology. She is particularly interested in knowledge production and the moments when knowledge claims are linked with the negotiation of standards or policies, but also situations that are more about caring and doing than knowing. She has conducted research on clinical trials and medical evidence, 'lay practices' and products for heart health and is currently researching antimicrobial resistance and inequality in the UK and the US. More info: https://profiles.sussex.ac.uk/p204943-catherine-will

Rosalind (Ros) Williams is a Lecturer in Digital Media and Society, and currently a Wellcome Trust Research Fellow in the Department of Sociological Studies at the University of Sheffield. Her research interests are in race, biomedicine, and the technologies, data and practices that connect them. Previous research has explored the use of digital (and non-digital) technologies in self-tracking practices, and the role of standards and racialised logics in biobanking. Currently, Ros is working on a project exploring how race is mobilised in the digital and face-to-face recruitment of donors to stem cell registries. More info: https://www.sheffield.ac.uk/socstudies/people/academic-staff/ros-williams

Liam Cole Young is an Associate Professor of Media Studies in the School of Journalism and Communication at Carleton University in Ottawa, Canada. He is the author of *List Cultures: Knowledge and Poetics from Mesopotamia to BuzzFeed* (Amsterdam University Press, 2017). He is working on a book that explores salt as an "elemental" medium of culture and civilization. His recent work has been published in *Theory, Culture and Society, Cultural Politics, Canadian Journal of Communication* and *Amodern*.

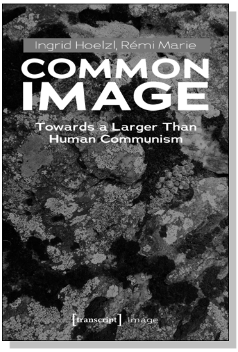
Malte Ziewitz is Assistant Professor at the Department of Science & Technology Studies at Cornell University. An ethnographer and sociologist, he studies the changing role of governance and regulation in, of, and through digitally networked environments. His recent work looks at the use of patient feedback in the British hospitals, the search engine optimization industry, and people's attempts to repair a sub-par credit score. At Cornell, he directs the Digital Due Process Clinic, a clinical research program that helps ordinary people to cope with, understand, and challenge automated decision systems.

Cultural Studies



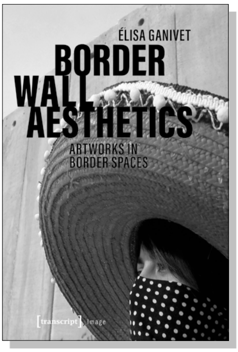
Gabriele Klein
Pina Bausch's Dance Theater
Company, Artistic Practices and Reception

2020, 440 p., pb., col. ill.
29,99 € (DE), 978-3-8376-5055-6
E-Book:
PDF: 29,99 € (DE), ISBN 978-3-8394-5055-0



Ingrid Hoelzl, Rémi Marie
Common Image
Towards a Larger Than Human Communism

2021, 156 p., pb., ill.
29,50 € (DE), 978-3-8376-5939-9
E-Book:
PDF: 26,99 € (DE), ISBN 978-3-8394-5939-3

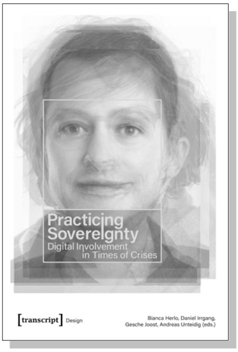


Elisa Ganivet
Border Wall Aesthetics
Artworks in Border Spaces

2019, 250 p., hardcover, ill.
79,99 € (DE), 978-3-8376-4777-8
E-Book:
PDF: 79,99 € (DE), ISBN 978-3-8394-4777-2

All print, e-book and open access versions of the titles in our list are available in our online shop www.transcript-publishing.com

Cultural Studies

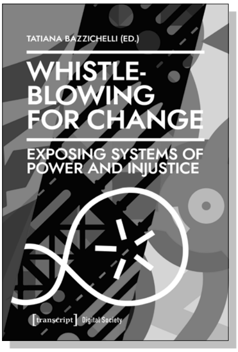


Bianca Herlo, Daniel Irrgang,
Gesche Joost, Andreas Unteidig (eds.)

Practicing Sovereignty

Digital Involvement in Times of Crises

January 2022, 430 p., pb., col. ill.
35,00 € (DE), 978-3-8376-5760-9
E-Book: available as free open access publication
PDF: ISBN 978-3-8394-5760-3

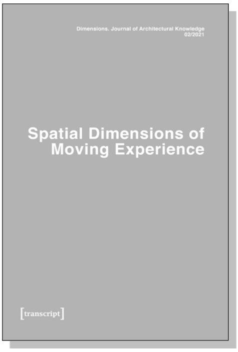


Tatiana Bazzichelli (ed.)

Whistleblowing for Change

Exposing Systems of Power and Injustice

2021, 376 p., pb., ill.
29,50 € (DE), 978-3-8376-5793-7
E-Book: available as free open access publication
PDF: ISBN 978-3-8394-5793-1
ISBN 978-3-7328-5793-7



Virginie Roy, Katharina Voigt (eds.)

Dimensions. Journal of Architectural Knowledge

Vol. 1, No. 2/2021:

Spatial Dimensions of Moving Experience

2021, 228 p., pb., ill.
39,00 € (DE), 978-3-8376-5831-6
E-Book: available as free open access publication
PDF: ISBN 978-3-8394-5831-0

All print, e-book and open access versions of the titles in our list are available in our online shop www.transcript-publishing.com