UNDERSTANDING DATA STUDIES

A METHODOLOGICAL AND CONCEPTUAL INQUIRY INTO RESEARCH ON DATAFICATION

A thesis submitted in
fulfilment of the requirements for
the degree of Dr. phil.

University of Bremen

by Irina Zakharova

Supervisors:
Andreas Breiter, University of Bremen
Anne Kaun, Södertörn University

Bremen, 31.03.2022

The following people served as readers for this thesis:

Prof. Dr. Andreas Breiter, University of Bremen

Prof. Dr. Anne Kaun, Södertörn University

Prof. Dr. Christian Katzenbach, University of Bremen

Abstract

Data are increasingly interwoven in various aspects of our social lives. The Covid-19 pandemic is an illustrative example of the role data play in our society: worldwide, researchers, public health authorities, media, and laypeople have been discussing epidemiological data and using them for decision-making. In our everyday lives, too, many kinds of data are produced. For instance, as we make online searches via search engines, chat with loved ones and friends via social media, or use a maps app on our smartphone to find our way in an unfamiliar place. Businesses, civic organisations, public and political actors all rely on data for their operations and decision-making. Regardless of the societal domain and area of application, data produced nowadays are increasingly digital. To use these digital data for decision-making, and literally anything else, people rely on computational technologies. With these, digital data can be processed, recombined, operated with, used, and sold. Going hand in hand with the pervasiveness of data in our society is datafication—"transforming all things under the sun into a data format" (van Dijck, 2017, p. 11).

Alongside with these societal transformations, researchers across manifold academic disciplines and fields from computer science and information systems research to sociology, media studies, communication research, humanities, and education research are working on topics concerning this *datafied* society. In the recent years, the body of work about the datafied societies and various implications of datafication processes has been consolidating under the terms 'critical data studies' or just 'data studies', drawing on various ontological, epistemological, theoretical, and methodological approaches to studying datafication processes. How, then, in this manifold of perspectives, academic knowledge about datafication processes and our datafied societies is produced? What is 'critical' in data studies? How do scholars conducting research on datafication reflect about "what matters we use to think other matters with; […] what stories we tell to tell other stories with" (Haraway, 2016, p. 12) in their studies? Especially considering that scholars also rely on digital data designed and produced for other purposes than academic studies in their research, these reflective questions become critical. The conceptual and methodological inquiry into empirical research on datafication, presented in this thesis, provides such a reflection of the emerging academic field of data studies. With my thesis, I advance our understanding of how what is known about datafication and datafied societies is produced.

Drawing on the concepts such as methods' performativity and methods assemblage (Law, 2004), practice theory, the body of work in sociology of knowledge, and on arguments from feminist research, I am particularly interested how methodological choices of researchers co-produce knowledge about datafication processes in our societies. The concept of methods assemblage allows such an investigation as it brings together the researchers with their personal and institutional positionings, the researched actors and things, the empirical site of practice where research is conducted, and particular research procedures. I take methods' performativity as a starting point of analysis rather than analytical challenge. With the literature analysis of current empirical studies on datafication and expert interviews with datafication scholars, my thesis makes several methodological and conceptual contributions to the literature in data studies and adjacent fields concerned with the datafied society. *Methodologically*, I map out methods assemblages of empirical datafication research and develop a heuristic for their analysis. Taken together, they can be used as a reflection tool for advancing sensitivities to the manifold of empirical phenomena addressed in research as 'datafication'. *Conceptually*, I show how empirical datafication research produces re-situated conceptualisations of datafication. Further, I discuss the role of critique in data studies and propose pathways for further, generative, care-ful critique, contributing to the literature bridging data studies with feminist traditions of thought. I believe, in taking performativity of methods in data studies as an analytical point of departure, I could further our understanding and practices of engaging with complexity of our datafied societies in a productive way.

Zusammenfassung

Daten spielen zunehmend eine wichtige Rolle in verschiedenen Lebensbereichen. Ein Beispiel dafür bietet die Covid-19 Pandemie. Weltweit haben Forschende, Gesundheitsexpertinnen und -experten sowie -behörden, Medienvertreterinnen und -vertreter, sowie die Öffentlichkeit über epidemiologischen Daten gesprochen und diese für Entscheidungsfindung verwendet. Im Alltag sind die Daten ebenfalls überall zu finden: Zum Beispiel, wenn wir eine Suchanfrage über eine Suchmaschine eingeben, mit unseren Freunden und Verwandten über soziale Medien kommunizieren, oder eine Karte über die App auf unserem Smartphone aufrufen, um unseren Weg an einem unbekannten Ort zu finden. Unternehmen, zivilgesellschaftliche Organisationen, öffentliche und politische Institutionen verwenden auch Daten in ihrer täglichen Arbeit. Vom gesellschaftlichen Lebensbereich und Anwendungskontext unabhängig, sind diese Daten heutzutage zunehmend digital. Um diese Daten zu nutzen, werden Informationstechnologien verwendet. Mit deren Hilfe können digitale Daten prozessiert, rekombiniert, verwendet, und verkauft werden. Mit dieser Durchdringung der Daten in alle gesellschaftlichen Lebensbereiche geht auch die Datafizierung einher: van Dijck (2017) beschreibt diese als eine Transformation aller Dinge der Welt in ein Datenformat (p. 11).

Neben diesen gesellschaftlichen Veränderungsprozessen beschäftigen sich auch Forschende in unterschiedlichen akademischen Disziplinen und Feldern von Informatik bis hin zu Soziologie, Medien- und Kommunikationswissenschaft und Bildungsforschung mit den Fragen rund um diese *datafizierte* Gesellschaft. In letzten Jahren hat sich Literatur rund um die datafizierte Gesellschaft unter dem Begriff „critical data studies“ oder auch „data studies“, als (kritische) Datenstudien konsolidiert, die sich unterschiedlicher ontologischer, epistemologischer, theoretischer und methodologischer Grundlagen und Ansätzen zur Erforschung der Datafizierungsprozesse bedienen. Vor dem Hintergrund dieser Vielfältigkeit stellen sich die Fragen, wie das Wissen über Datafizierung der Gesellschaft produziert wird, was ‚kritisch' an Datenstudien ist und wie Forschende in diesem Bereich über ihre Projekte reflektieren. Außerdem werden die Forschenden selbst zu Teilnehmenden an Datafizierungsprozessen, beispielsweise wenn sie für ihre Studien digitale Daten verwenden, welche von kommerziellen Unternehmen für andere Zwecke als Forschung kreiert werden. Die hier vorgestellte konzeptuelle und methodologische Untersuchung der empirischen Forschung zu Datafizierung zielt darauf ab, eine solche Reflektion der hier skizzierten Datenstudien vorzunehmen. Mit meiner Studie wird ein besseres Verständnis davon erreicht, wie unser Wissen über Datafizierung produziert wird.

Auf den Konzepten der Methodenperformativität und Methodenassemblagen (Law, 2004) aufbauend, sowie unter Berücksichtigung der praxistheoretischen, wissenssoziologischen und feministischen Argumente untersuche ich, wie verschiedene methodologische Ansätze das Wissen über Datafizierung der Gesellschaft koproduzieren. Das Konzept der Methodenassemblagen erlaubt eine solche Untersuchung, da es die Forschenden und ihre persönlichen sowie institutionellen Positionierungen, die Forschungsobjekte, den empirischen Kontext einer jeden Studie, sowie die spezifischen Forschungstätigkeiten analytisch unter einem Dach zusammenbringt. Dabei positioniere ich Methodenperformativität als analytischen Ausgangspunkt meiner Untersuchung. Durch eine Literaturanalyse aktueller empirischen Datafizierungsforschung und durch Experteninterviews mit Forschenden auf diesem Bereich mache ich methodologische und konzeptuelle Beiträge zum Forschungsfeld der Datenstudien sowie den anderen Disziplinen, die sich an Datenstudien angrenzend ebenfalls für datafizierte Gesellschaft interessieren. *Methodologisch* kartiere ich Methodenassemblagen empirischer Datafizierungsforschung und erarbeite eine Heuristik für deren Analyse. Zusammen bilden diese ein Reflektionswerkzeug mit dessen Hilfe Sensibilität gegenüber vielfältigen empirischen Phänomenen, die als ‚Datafizierung' bezeichnet werden, weiterentwickelt werden kann. *Konzeptuell* erläutere ich, wie in empirischer Datafizierungsforschung re-situierte Konzeptionen der Datafizierung entstehen. Zudem diskutiere

ich die Rolle der Kritik in Datenstudien und mache einen Vorschlag, wie der Weg für die Weiterentwicklung dieser zu einer generativen, sorgsamen Kritik aussehen könnte. Dabei mache ich einen Beitrag zu Literatur, die Datenstudien und feministische Forschung zusammenbringt. Ich bin überzeugt, dass ich mit meiner Untersuchung, die Methodenperformativität in Datenstudien als ihren analytischen Ausgangspunkt betrachtet, unser Verständnis über und Praktiken in der Auseinandersetzung mit den komplexen Prozessen in einer datafizierten Gesellschaft voranbringen konnte.

Acknowledgements

Like any other academic book, my thesis would have not been possible without many people who accompanied me on my PhD journey. First, I would like to express my gratitude to my supervisor Prof. Dr. Andreas Breiter. Welcoming me in a new position as a PhD candidate at the University of Bremen and a researcher in the Institute for Information Management Bremen (ifib), Andreas provided me steady support as I navigated my new disciplinary affiliation and a new career path. Providing an atmosphere of encouragement and academic freedom and creating opportunities for new ideas and tasks, Andreas contributed a great deal to my professional and personal growth, which I am deeply thankful for.

I appreciate very much Prof. Dr. Anne Kaun and Prof. Dr. Christian Katzenbach also joining my PhD project as a second supervisor and an examiner, respectively. I would like to thank Anne for her always encouraging feedback and important advice on improving my dissertation.

I would like to thank all other members of my committee for their support and their interest in my research: Prof. Dr. Andreas Hepp, also for participating in my committees twice—first, for my master's thesis and second, for my PhD, Dr. Juliane Jarke, and Yan Brick. I also extend my deepest gratitude to all the experts who participated in my empirical research during the Covid-19 pandemic, despite all the constraints. I am grateful for your interest in my work, the openness you all showed in the interviews, and the words of encouragement.

I am indebted to Dr. Juliane Jarke for her impact on my research. As a master's student at the University of Bremen, I attended multiple courses Juliane has held. Through these courses and through her mentoring, Juliane not only introduced me to the intriguing research domain of Science and Technology Studies, but also to academia as a career path. I am deeply grateful for both. I also thank Juliane for her mentorship and for the joy of thinking and writing together that you radiate and taught me to appreciate so much.

I would like to extend my sincere gratitude to all my colleagues, past and present, at ifib and the Research Group Information Management (AGIM) at the University of Bremen, notably Dr. Hendrik Heuer, Tjark Raabe, Angelina Lange, Izumi Klockmann, Gabriela Carolina Molina León, Dr. Philipp Krieter, Dr. Emese Stauke, Dr. Anja Zeising, Dr. Marion Brüggemann, Julia Finken, Dr. Stefan Welling, Ulrike Gerhard, and also Dr. Dafna Burema. I also would like to thank all my colleagues from the DATAFIED research project for their continuous support, engaging discussions, and learning experience. My special thanks go to the colleagues in the Department for Media and Communication Studies of the Södertörn University, which I had the pleasure of visiting in the last months of my PhD project. Thank you all for welcoming me, your interest in my research, fruitful discussions, and your advice. I also thank my friends from the "blue flow" ECREA 2020 PhD summer school.

I also wish to thank my friends Dr. Chen Weiyi, Tatiana Damyanova and Illian Damyanov, Britta Hermann, Julian Schierenbeck and Maria Schierenbeck, Lea Telle, Yakov Tokarev, Julia Willenbrock, Melanie Wrobel. My eternal gratitude goes to my family—Tatiana Yurievna Zakharova, Anatoly Illich Zakharov, Nadira Zheferovna Abrasulina—whose support and love never ceased to reach me across distance and time apart. I also sincerely thank the family I found in Bremen—Sebastian Kettler, Silvia Kettler, Sophia Kettler. Sebastian, thank you so much for supporting me all the way on this journey! Finally, these acknowledgements would not be complete if I did not mention our two cats, who were, literally, around me most of the time I was writing this thesis during the two years of the pandemic. Rather than thanks they would, however, appreciate (and get) some treats.

To my parents

List of figures

List of tables

1 Introduction

Data are increasingly interwoven in various aspects of our social lives. The Covid-19 pandemic, during which the most of this thesis was written, is an illustrative example of the role data play in our lives: worldwide, researchers, public health authorities, media, and laypeople have been discussing virus loads, virus reproduction values, quantity of available hospital beds, or number of administered vaccinations for two years to the time of writing. Based on these and other data, policy-makers made decisions about pandemic prevention measures. During the war escalation on the territory of Ukraine raging as I write this, social media content became an important source of communication and information about the developments in the affected regions not at least for journalistic work and media coverage. Addressing another pressing crisis, a new climate report was published shortly before I finished this manuscript: for that climate researchers collected evidence about and developed strategies of dealing with global climate change based on chemical, biological, geological data, data from social sciences and other research disciplines (see NASA, n.d.). In reporting about and/or tackling these crises, various kinds of data are used by researchers, political actors, journalists, civic society, and laypeople for decision-making, communication, and visualisation of information. Some of these data can be produced with specific sensory devices (for example for measuring carbon dioxide in the air), other data, for example about the vaccines administered during the Covid-19 pandemic, are produced by people responsible for documenting vaccination processes in each country. Other kinds of data are being produced in our everyday lives, for instance as we make online searches via search engines, chat with loved ones and friends via social media, or use a maps app on our smartphone to find our way in an unfamiliar place. Regardless of the societal domain and area of application, data produced nowadays are increasingly digital—generated with the help of computer code and stored in Bits and Bytes. To use these digital data for decision-making, and literally anything else, people rely on computational technologies. Data are being processed with the help of algorithms—techniques which allow turning certain input data in an output data with the help of computational rules, based on mathematics and statistics. These algorithms are run via computer programs—software—which, following Berry (2015 [2011]) is "a tangle, a knot, which ties together the physical and the ephemeral" (p. 3). Such software, however, exists not outside of other societal processes, it is deeply embedded in the organisational, material, political, and economic processes relevant for the software providers. To account for these intertwining socio-technical relations, the term 'information systems' is used that stresses how data processing via such technological and computational means serves for *informing* various actors' actions and applying the knowledge gained from the data practically in accordance with their various goals. For example, some activist communities use community-gathered digital data for their work and in their attempts to change and improve certain policies: such projects are discussed by D'Ignazio and Klein (2020)in their book "Data Feminism", notably the Our Data Bodies project, directed at supporting communities in issues of data collection and human rights (https://www.odbproject.org; for outcomes of the project supporting sharing knowledge about data collection with various communities see Lewis et al., 2018). Governments, political, and public actors also exhibit growing interest in relying on digital data, for example for decisions about public services provision (*Der AMS-Algorithmus*, n.d.), surveillance (Koble, 2019), policing (Metz & Satariano, 2020), or for tracing the spread of the Covid-19 virus during the pandemic in collaboration with Google and Apple (Apple, n.d.; Google, n.d.; European Commission, n.d.).

Going hand in hand with the pervasiveness of data is datafication—"transforming all things under the sun into a data format" (van Dijck, 2017, p. 11). The processes of datafication are as manifold as are the diverse ways to process, operate and work with, use, regulate, delete, ignore, or resist digital data. Datafication here refers to various empirical phenomena in which digital data

and means of their production and use are central. For example, in 2020 datafication of education was widely discussed in the United Kingdom after the results of A-level exams were published and revealed how due to the grading algorithm's design many exam results were downgraded, leading to its massive critique (Adams, 2020; Burgess, 2020). 2021 there have been a series of whistle-blowing and journalistic reports about social media sites such as Facebook or TikTok which illustrated for broader publics how these companies gather, operate with, and profit from data generated by their users (e.g. n.a., 2021; Paul & Anguiano, 2021). Figure 1-1 shows few newspaper headlines from around the world addressing the role of data in our societies in the last ten years and the ways in which digital data are being used, resisted against, and regulated. For example, an article titled "The Age of Big Data" (Lohr, 2012) was published 2012 in The New York Times discussing how various businesses and public institutions were embracing data science—data analysis techniques drawing on methods and concepts from computer science, statistics, and mathematics—for identification of new patterns in data and development of new knowledge that could be applied for their organisational goals. The article "Deutsche fürchten Big Data" (German: Germans fear big data) (Adam, 2016) in the Süddeutsche Zeitung highlights the perceived privacy risks of digital data. An article "Big data is going to shape our future cities. Will it treat us all equally?" (Bogle, 2018) published by ABC Science highlights multiple kinds of data produced in modern cities and how these can be analysed more or less successfully for different purposes, ranging from urban planning to crime prediction.

The headlines mentioned here and shown in figure 1-1 use a variety of terms such as 'data', 'big data' (where 'big' refers to the volume of datasets used for further analysis), 'algorithms' and 'artificial intelligence' (algorithmic techniques based not on rather simple computational rules, but on procedures through which algorithms are trained and 'learn' according to certain criteria based on existing sets of data). Other terms, also often mentioned in relation to data and datafication, are software, (social media) platforms, and apps. For example, covering the role of TikTok in reporting about the next leap of military escalation during the war in Ukraine, an article "War as seen on TikTok: Ukraine clips get views whether true or not" (Milmo & Farah, 2022) published in the Guardian discusses the role of the platform's algorithms in the rapid spread of information from and about the zone of military conflict. While datafication and not algorithms or AI is the topic of my thesis, figure 1-1 and examples mentioned so far show how in public discussions, 'datafication' of various aspects of social lives is addressed through multiple terms such as algorithms, big data, or AI, while these terms are often used interchangeably. The abundance of recent and historical examples and headlines referring to various datafication phenomena demonstrates not only great public interest in the topic, but also the variety and range of implications datafication has on our social lives. So, the examples mentioned here touch upon issues of sustainability—analysed based on climate data, equality and equity—for instance in data about access to public health, mis(representation), bias—exemplified through the massive critique on the A-levels grading algorithms, economy—digital service providers' practices of data extraction and reselling, power and agency—addressing diverging possibilities different social actors have in control over the data, and related legal issues—for instance data protection and privacy of people using various digital services. These issues are among the central ones in social research and have been explored by social researches long before datafication become pervasive in our societies. The headlines and examples mentioned here, however, showcase how datafication adds additional layers of complexity, drawing attention to the socio-technical relationships and technological developments relevant to the understanding of these issues. Social research, therefore, is turning its attention on the processes of datafication as an important part of its body of work. In my thesis, I am interested in this work on datafication and discuss datafication predominantly from the perspective of social sciences.

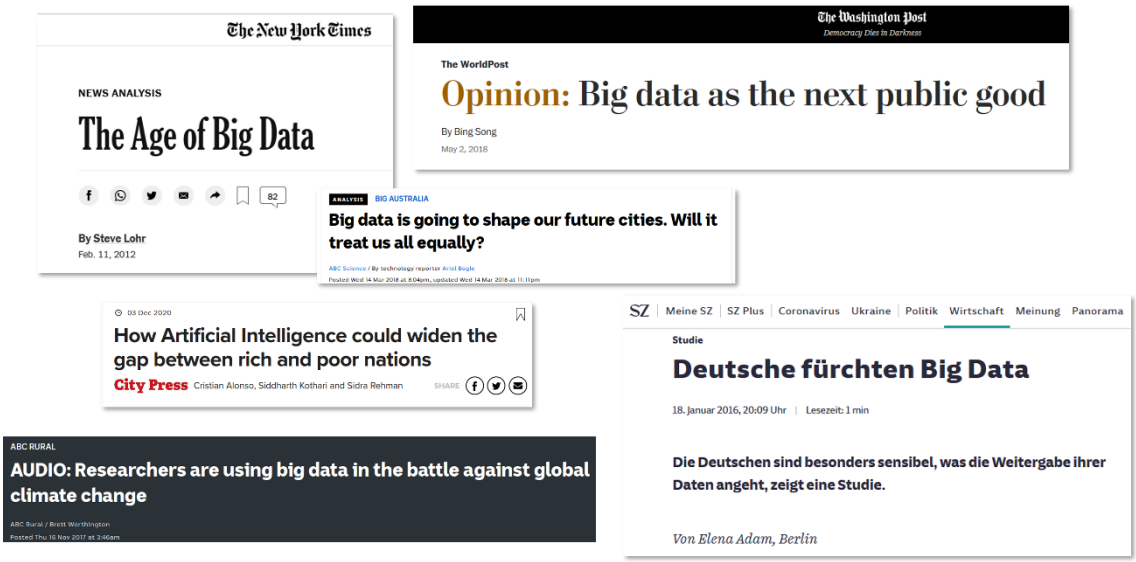

Figure 1-1 A sample of newspaper and magazine headlines about big data, datafication, and AI

Researchers in social sciences attend to datafication as one of the current and academically relevant topics. For example, literature reviews conducted by Kennedy and colleagues (2020) and by Flensburg and Lomborg (2021) showcase both the amount of emerging conceptual and empirical research on the topic of datafication, and the themes and concerns this research covers. According to Flensburg and Lomborg (2021), authors from more than 18 different research domains and fields ranging from communication, education, sociology, computer sciences to law, health, geography, philosophy, and economy, among others, have published about datafication. Central to this research are critical perspectives on datafication opposing initial, enthusiastic literature about digital data (addressed in academic discourse in early 2010s as 'big data'). Such literature claimed that through the technological possibilities of gathering big amounts of data, speedy processing, and recombination of these data for finding patterns in the datasets new opportunities for 'evidence-based', 'data-driven' knowledge production open (e.g. Mayer-Schönberger & Cukier, 2013). Scholars taking a critical position on digital data and datafication dismiss such enthusiasm as a positivist view on the complexities in which digital data relate to societies. These critical scholars conceptualise data as not 'raw' or neutral, but rather processed by other actors, for example technology providers, according to their goals (Borgman, 2015; boyd & Crawford, 2012; Gitelman, 2013). So, common for critical perspective on digital data are notions of their relationality and recursivity: digital data not only represent phenomena, but also reconfigure processes and objects they are meant to represent. Datafication, then, reflects such continuous reconfigurations of data and phenomena they are meant to represent and, accordingly, is understood as on ongoing socio-technical process of translating social practices and phenomena into data (e.g. Couldry & Hepp, 2017; Jarke & Breiter, 2019; van Dijck, 2014; Williamson, 2017).

Some scholars, however, have been critiquing various aspects of datafication research and related fields on which it extends (such as algorithm studies, software studies, digital sociology). Some strains of critique primarily address the issues explored in the body of work addressed as critical data studies: an interdisciplinary body of work, encompassing research on the manifold of datafication processes in our—*datafied*—society from different fields and disciplines within and beyond social sciences: media and cultural studies, communication research, education research, sociology, data and computer science, information systems research, to name only a few. For example, scholars conducting audience research in media and communication studies point out that individual users are all too often neglected in data studies as opposed to the analytical attention to the big proprietary technology providers (big tech) (e.g. Dencik, 2020; Livingstone, 2019; Mathieu

& Pruulmann Vengerfeldt, 2020). Others call for investigating the everyday, mundane situations in which laypeople are affected by datafication processes and re-enact these on a daily basis as a part of their routines (e.g. Kennedy et al., 2020, 2021). Another strain of critique argues that concepts and theoretical approaches developed in other disciplines such as the notion of 'care' are helpful in reconceptualising datafication research und studies of data, including data science, in order to acknowledge and reflect on the normative, ethical, and affective relations between digital data and other human or non-human actors in the society (e.g. Fotopoulou, 2020; Lupton, 2020; Zakharova & Jarke, 2022; Zegura et al., 2018). In my thesis, I build on this critical body of work in critical data studies. I consider research on datafication a part of this academic literature. With my thesis, I contribute to critical data studies by advancing our understanding of academic knowledge production in these studies methodologically and conceptually.

While critical data scholars offer critique of datafication processes, they face a methodological challenge: in order to explore, understand, and explain datafication, empirical social research also relies on different kinds of digital data. Conducting empirical studies on datafication processes, scholars might be interested in investigating relevant kinds of digital data or ways in which these data are being processed and used. Some scholars outline methodological challenges for studying datafication processes. So, Tufekci (2014a), discusses pitfalls of analysing data from social media, and particularly Twitter, in regard to the generalisability and ethics of such studies. Others point out that social research sometimes relies on data collection instruments developed by non-academic actors for their commercial and other purposes: for example, using likes and reposts on social media as research sampling criteria (e.g. Lindgren, 2020; Rieder & Röhle, 2017). Further, researchers' lives are also datafied, as they participate in various societal activities professionally and personally: for instance, searching for new publications online, using information systems to store and analyse empirical research material, or presenting their research on social media such as Twitter or ResearchGate.

Another line of methodological critique touches upon the very foundation of critical perspective on datafication. So, Law (2004) argues that exploring relational, recursive processes—the description of processes enclosing datafication—is methodologically challenging: "while standard methods are often extremely good at what they do, they are badly adapted to the study of the ephemeral, the indefinite and the irregular" (p. 4). Methods, here, can be understood as various techniques for collection and analysis of research material. Responding to this, various methodological approaches with the focus on the research on digital data are emerging in social sciences that specifically address processuality of datafication by studying relations developing in time and space or practices of negotiation (e.g. Decuypere, 2021; Ratner & Gad, 2018; Velkova, 2018). At the same time, special issues' and handbooks' authors and editors reframe methods from mere 'tools' of conducting empirical research to an object of inquiry: they question and explore, how and which methods are suited for study of which datafication processes, what do these methods help to learn about datafication, and what methodological changes research undergoes when it is directed at datafication processes (Hand & Hillyard, 2014; Kara, 2020; "Problematizing Methodological Simplicity in Qualitative Research: Editors' Introduction" by Koro-Ljungberg & Mazzei, 2012; Kubitschko & Kaun, 2016; "Methods for datafication, datafication of methods: Introduction to the Special Issue" by Lomborg, Dencik, et al., 2020; Lury & Wakeford, 2012; Marres, 2017; Snee et al., 2016; Woodward, 2019). In sum, while datafication researchers aim to conceptualise datafication processes critically and study digital data and/or their implications, these scholars at the same time participate in other datafication processes themselves. What does it mean, then, for our understanding of datafication and how can this understanding be advanced?

In their introduction to a special issue in the International Journal of Communication, Hepp and colleagues (2018) call for putting digital data in context (see also Breiter & Hepp, 2018). The editors argue that such a contextualisation of digital data can take place alongside three dimensions: that of the relevant academic discourses, the methods applied, and the empirical fields

studied (Hepp et al., 2018, p. 440). In my doctoral thesis, I follow this call for contextualising data—and datafication processes related to these—by focusing on methods and methodologies applied to study them. For example, Kennedy and colleagues (2020) point out that data studies and datafication research lack thorough documentation and reflection of research methods in reporting results (p.44). The well-established body of work on sociology of scientific knowledge (e.g. Collins & Pinch, 2012; Knorr-Cetina, 2002; Latour & Woolgar, 1986; Pinch & Bijker, 1984; Sismondo, 2012) which I draw upon here, has long been concerned with the practical, epistemological, and methodological issues of academic knowledge production, including the role of informal relations, implicit knowledge, institutional and material, physical contexts of research, and also practices of reflection. So, in regard to the methodological critique and challenges of data studies, an argument made by Donna Haraway (2016) on the importance of reflection in academic knowledge production becomes particularly pressing:

> "It matters what matters we use to think other matters with; it matters what stories we tell to tell other stories with; it matters what knots knot knots, what thoughts think thoughts, what descriptions describe descriptions, what ties tie ties. It matters what stories make worlds, what worlds make stories" (p. 12).

How do datafication scholarship and critical data studies reflect on the matters they use 'to think other matters with'? How do they reflect on the stories they tell 'to tell other stories with'? How can such reflection be established and how can the notion of 'critique' in *critical* data studies be directed at the data studies themselves? Most recently, there have been calls to attend to theoretical eclectics to gain a deeper understanding of the theories underlying (digital, computational) methods used to study datafication. For example, Lindgren (2020, p. 13) and Rieder and Röhle (2017, p. 123) argue that researchers relying on computational methods should have (more) theoretical sensitivity to the histories and development of the respective methods. This sensitivity requires not only a deep knowledge in both applied theories and methods, but also ways to make the gained results intelligible for others. Complementary calls in the methodological literature point to the central role of methodological sensitivity and reflexivity in the academic knowledge production (e.g. Springgay & Truman, 2018). In recent literature, a notion of 'care' could also be found in calls for more careful and reflexive academic research (e.g. Law, 2021). I follow this call and discuss in conclusion of my thesis, how the contributions made in my thesis open new, critically generative pathways for the further development of data studies as an academic research field.

With my thesis, I contribute to this literature by conducting an inquiry into current empirical datafication research in social sciences in order to map out research methodologies applied for studying datafication empirically and to explore how these methodologies are reflected upon by datafication scholars. I understand methodologiess as complex and situated assemblages of human and more-than-human elements held together through research practices (see e.g. Law, 2004; Mol, 2002) and refer to them as methods assemblages. It has also been widely accepted that methods assemblages are performative and co-produce the phenomena that they are applied to study (e.g. Barad, 2007; Law & Ruppert, 2013; Savage, 2013). Paraphrasing Law and Ruppert (2013, p. 234), the question motivating my study was, what is datafication according to the current empirical research on it and how is datafication conceptualised through and within methods assemblages. Based on the notion of methods' performativity, I argue that methods assemblages and concepts of datafication need to be explored in concert with one another (Zakharova, 2021). Therefore, I am interested in the ways different methods assemblages contribute to the knowledge production in critical data studies and what kinds of concepts about datafication processes they produce. As datafication becomes more and more important for greater number of societal issues, it also means that researchers across different research disciplines will continue to turn their increasing attention to datafication processes, data and their implications for various actors and societal domains. Thus, it is ever more important to investigate how various methods assemblages,

embedded in different academic disciplines, co-produce various kinds of concepts about datafication.

My thesis primarily aims to contribute to data studies by mapping out how datafication processes are *assembled* in data studies methodologically and conceptually. With that, my thesis contributes to critical data studies in three ways. First, I provide a reflection of the current empirical body of work by exploring methods assemblages datafication scholars enact in their research projects. Second, I conceptualise currently applied concepts about datafication arising from such empirical research and put these in relation to the identified methods assemblages. Further, I advance the understanding of how data studies produce knowledge about datafication, how data scholars reflect on their research practices, and what methodological sensitivities such reflection advances. Together, these contributions present a heuristic for reflecting on empirical research on datafication. As the body of work in data studies comprises research from multiple disciplines and fields, notably media and cultural studies, algorithm studies, digital sociology, and communication research, my methodological and conceptual inquiry also makes contributions to studies on datafication within these fields. Finally, I make a contribution to methodological literature by exploring relations between methods assemblages, digital data, society, and academic knowledge production. These contributions are primarily conceptual and methodological. Nevertheless, heuristic for methodological and conceptual reflection constructed here provides datafication researchers across different academic disciplines as well as practitioners conducting studies about datafication processes with a practical reflexive tool advancing their sensitivity required for rigorous and care-ful engagement with datafication processes.

I provide such a contribution by attending to the following research questions (RQs).

- RQ 1: Which research methods are used in what ways in social sciences to study datafication processes empirically?
- RQ 2: How the concepts about datafication determine the choice and use of research methods? How these methods re-produce these conceptions by re-situating datafication processes empirically?
- RQ3: How can datafication scholars reflect on the use of different methods applied for studying datafication processes empirically and what kinds of methodological sensitivity such reflection requires?

To achieve this goal, I apply a practice-theoretical approach (Schatzki, 2002; Nicolini, 2009b) for my inquiry. With that, I conceive of methods assemblages applied by datafication scholars as methods-assemblages-in-practice (Law, 2004) as they bring together various actors involved in academic research. A practice-theoretical approach developed by Schatzki (2002) helps to understand methods assemblages: "doings" like empirical research practice and "sayings" (p. 72), for instance in spoken and written reports about this research. Empirically, methods assemblages, thus, can be understood as related practices of data collection, data analysis, reflection, and reporting that make datafication intelligible to others. I follow an iterative approach to my research design, drawing on Nicolini's (2009b) concept of zooming in and zooming out: it allows studying practices that hold together a methods assemblage and, relating these to the broader context of a respective academic discipline, academic knowledge production, and empirical site of practice in which the researched datafication processes are enacted. I synthesised a literature analysis of datafication scholarship with expert interviews with authors of the sampled publications and further datafication researchers. In the interviews, experts gave accounts of their research practices reported in academic writing and reflected on their own methodological choices and approaches to studying datafication processes empirically. These reflections situate findings from my literature synthesis in various contexts of academic scholarship. Results discussed in the following chapters of my thesis present anonymised, aggregated results of my inquiry. While my focus in the analysis of both sampled literatures and expert interviews was on research methodologies and concepts about

datafication, I developed inductive categories throughout my analysis. These categories, ultimately, present various methods assemblages identified in my sample, different concepts about datafication produced with these methods assemblages, as well as specific research methods—techniques of data collection and analysis applied, issues related to the research politics and academic knowledge production (e.g. writing and publishing practice)—both common to social sciences in general and to datafication research in particular. Finally, a category addressing researchers themselves and their subjectivities completes my analytical framework. With the help of this framework, three distinct methods assemblages were constructed which reflect the sample of literatures and interviews. These methods assemblages are enacted for 1) exploring encounters with data and their representations such as visualisations, 2) tracing dynamics of data infrastructures and data movement, 3) reconstructing datafied regimes. Based on other categories created inductively, I construct a heuristic for analysing and reflecting on these methods assemblages.

In the following, I outline the structure of my thesis. To provide an overview over existing concepts about datafication in chapter 2, I begin with a literature overview of the current datafication research in social sciences and other research disciplines. For example, I briefly review concepts from computer sciences and information systems research required to understand technological underpinnings of digital data. In this chapter, I recount how critical data studies developed in the past ten years (in particular after the publication of the book by Mayer-Schönberger & Cukier, 2013), what understandings of digital data have been proposed and critiqued since, and how these data relate to datafication processes. I also review the multiplicity of theoretical concepts about datafication and situate these within other academic discourses in social sciences. The core part of this chapter is dedicated to the emerging field of data studies. With this chapter, I outline empirical research on datafication as a research object of my thesis and situate it as a part of the broader data studies scholarship.

Chapter 3 reviews current methodological discourses about empirical examinations of digital data and datafication processes and situates these historically as well as within scholarly work on academic knowledge production. This brief historical overview is by no means exhaustive and serves two main goals. First, academic research can be conceptualised according to the communities of scholars (Waisbord, 2019), established traditions of thought (Kuhn, 2020 [1976]), and situated research practices (Knorr-Cetina, 2002): in this chapter, I recount how these evolved around the use of specific research techniques/methods and related philosophical, ontological, epistemological perspectives. This is relevant for my empirical investigation, in which I engage with empirical studies of scholars representing diverse research and methodological communities. Furthermore, to be able to understand and analyse a broad range of methods, I need to build up an understanding of a broad variety of methodologies and underlying paradigms. Second, by situating methodological discourses historically, I review how these methods are changing in relation to datafication. I also attend to the role of research methods in academic knowledge production by reviewing the 'double social lives of methods' debate (Law et al., 2011; Law & Ruppert, 2013; Ruppert et al., 2013) in relation to data studies and discuss the concept of methods' performativity. Overall, chapter 3 builds a conceptual foundation for constructing methods assemblages empirically in my qualitative analysis.

In the next chapter 4, I discuss the research design of my thesis. I begin by outlining practice-theoretical approaches as a methodological background of my thesis and reflect on their role for the outcomes of my research. Theories of practice provide helpful methodological tools to study the 'methods-assemblages-in-practice'. In the core of my research design is a literature synthesis consisting of an explorative quantitative and an inductive qualitative analysis of sampled academic articles combined with inductive qualitative analysis of expert interviews with datafication scholars. The chapter provides details about the literature sample and the applied techniques of data collection and analysis. Furthermore, I reflect on my research design, on conducting research and writing a PhD thesis during a pandemic, and how all of these shape the results and contributions of

my thesis. Overall, chapter 4 explains how a chosen combination of analytical concepts and research techniques allows reaching my goal of mapping out the field of data studies methodologically and conceptually.

In the following chapter 5, I discuss the results of my literature analysis. This chapter includes visualisations of the sampled literatures created with VOSviewer (van Eck & Waltman, 2007) as part of an explorative analysis. This quantitative exploration provides entry points for further qualitative analysis and the chapter is structured according to the quantitative findings. Further, I explore how authors of the sampled literatures situate datafication processes conceptually and then empirically, in relation to methods assemblages they enact and elements they assemble. Chapter 5 synthesises current empirical datafication scholarship, maps out methods assemblages enacted within the sampled literatures, and provides first insights into the empirically informed, situated concepts about datafication processes researchers develop. Drawing on results presented in this chapter, I develop a heuristic for describing and analysing methods assemblages.

Chapter 6 iteratively builds on the findings from the literature analysis and completes my literature synthesis. In this chapter I introduce each of the three constructed methods assemblages. These methods assemblages are situated in the ontological research and data politics discussed by interviewed experts. Findings presented in this chapter also serve as an addition to the heuristic illustrated in chapter 5, explicating different concepts about datafication processes, situated in the socio-technical relations and practices within which these datafication processes are enacted.

In chapter 7, I complete the heuristic for analysing methods assemblages and discuss what is at stake in data studies, drawing on Latour's (2004) famous critique of critical research. This chapter combines this analysis of critique in academic knowledge production with methodological sensitivities established in feminist traditions of thought (using the concept of care) and with the heuristic I constructed. I argue that critical reflection on methods assemblages and new, generative kinds of critical questions in data studies can be developed. Expanding on the notion of care-ful research (e.g. Law, 2021; Law & Lin, 2020) I discuss sets of sensitivities it can contribute to methodological reflection in data studies. In conclusion to my thesis, therefore, I plea for more care-ful data studies, attuned to empirical, conceptual, and ontological multiplicities (Mol, 2002) of datafication processes and open to thorough reflection and (self-)critique.

2 Digital data and datafication

For Christine Borgman, "'[d]ata' is the most elusive term of all" (Borgman, 2019, p. 2). The myriad ways in which digital data are generated, processed, analysed, stored, deleted, and broken—performed in societies by different actors, human and non-human ones, can be broadly defined as 'datafication'. Digital data, therefore, are the building block of datafication processes. In the second chapter of my thesis, I explore the elusiveness of data and discuss various ways to define them like 'digital traces' or 'big data', the manifold stories these notions tell us, and how understanding these notions leads us towards a better understanding of datafication processes. In this and the following chapters, the term 'data' is used in plural. In this chapter and in most cases in the following ones, the term 'data' is also used not in a colloquial meaning of the word, addressing, for example, empirical research material as 'data'—such colloquial use of the term can be found in empirical chapters and is made explicit in the writing. Rather, in all other cases I use the terms 'data' and 'digital data' synonymously to identify the multiple kinds of digitally produced, processed, and stored in various sociotechnical systems Bits and Bytes, addressed in datafication scholarship.

While data can be considered as a building block of datafication processes, in research, the increase of attention to such processes can be traced back to the seminal book by Mayer-Schönberger and Cukier (2013). The word 'datafication', however, has also been used earlier: for example, Brown & Duguid (2000) also use this term in a sense echoing the current critical discourses on datafication:

> "So it's not surprising that infoenthusiasts exult in the simple volume of information that technology now makes available. They count the bits, bytes, and packets enthusiastically. They cheer the disaggregation of knowledge into data (and provide a new word—*datafication*—to describe it)" (p. 12, original emphasis).

Drawing on this quote here, I begin the discussion on datafication and the role of digital data in our societies by first attending to relations between data, information, and knowledge. In this chapter, I review various discourses in the social sciences, first, in relation to data, in sections 2.1 and 2.2, and then in relation to datafication processes, in section 2.3. In section 2.4 I review how social sciences have approached data and datafication processes methodologically, with that turning to the core topic of my thesis. In section 2.5 of this chapter, I briefly introduce an emerging body of scholarship in datafication research and critical data studies as a core resource both for my conceptual reflection and my methodological investigation. Keeping in mind the question of methodological approaches to studying data, I review critical data studies literature in relation to other domains and fields such as digital humanities or computational social science. I conclude this chapter with a discussion on the role of 'critique' in academic knowledge production within critical data studies, which I come back to later in chapter 7. Addressing issues of academic knowledge production and critique, I set the direction for my research design and analysis of my empirical work, discussed in chapters 4 and 5-7, respectively.

2.1 About data

The term 'data' has multiple definitions stemming from various research disciplines and fields. In this section, I provide several perspectives in order to contextualise the relation between 'data' and other, closely related terms such as 'information' and 'knowledge'. Understanding this relation is helpful for the argument and contribution my thesis makes: 'data' in different senses of the term

can be considered as an important step in academic knowledge production. The goal of such knowledge production in times of "normal science" between what Kuhn (2020 [1976]) calls scientific revolutions is in solving "puzzles" to further and make existing knowledge more specific. Data, in the colloquial sense of the term in academic research, refers to the research material required to advance academic knowledge. As Krippendorff (1970) argues, methodologically, different disciplines rely on different kinds of such research data/material. Drawing on Coombs' theory of data, Krippendorff defines data as formalised research observations, while this formalisation takes place through e.g. measurement or recording. Such data as research material can be understood through following characteristics: data are durable, data as human-made artifacts, analysable, demonstratively reliable representations of something other than their own materiality, provide enough information to settle research questions (Krippendorff, 2016). In this view, data are mainly used to advance knowledge production in research through such analytical operations as drawing distinctions, aggregating, comparing, etc. An analysis drawing on these and other operations and techniques, according to Krippendorff (2016), needs to "preserve the information about the phenomena of interest to the extent relevant in selecting the answers to given research questions" (p. 488). This brief discussion of the term data as research materials used for academic knowledge production points to the relations between data, information, and knowledge. What happens to these relations, then, when digital data become research data? In order to understand that, first, all these terms require further contextualisation before we can proceed with a discussion of digital data *both* as an *element of empirical phenomena* studied by datafication scholars and as *research material for their inquiries*. In the remainder of this section, I turn to various perspectives on digital data in information science, information systems research, and computer science., followed by conceptions of data in social sciences.

Information science is centrally concerned with problems of knowledge representation, storage, search, and retrieval—e.g. in application forms of search engines, archives, and libraries (Stock & Stock, 2013). As this list suggests, in information science, the term 'information' is tightly related to 'knowledge' and draws on the concept of information as presented in Shannon's model of communication (e.g. Shannon & Weaver, 1949). This concept is widely adopted across various research disciplines both in information science and within social sciences. This model of communication puts forward transmission of information through a channel in form of a signal from a source to a receiver, while the source and receiver respectively encode and decode the signals. This model also points to the challenge of reproducing not only the signal, by the meaning this signal should bear through such a transmission. Regarding the role of data, Stock and Stock (2013) note in their Handbook of Information Science following.

> "Signals transmit signs. In semiotics (which is the science of signs), signs can be observed more closely from three points of view:
>
> — in their relations to each other (syntax),
>
> — in the relations between signs and the objects they describe (semantics),
>
> — in the relations between signs and their users (pragmatics).
>
> Shannon's information theory only takes into consideration the syntactic aspect. From an information science perspective, we will call this aspect data. Correspondingly, the processing of data concerns the syntactical level of signs, which are analyzed with regard to their type (e.g. alphanumerical signs) or their structure (e.g. entry in a field named "100"), among other aspects" (p.22).

Following this understanding from information science, the meaning and usage of such signs can be understood through the concepts of 'knowledge' and 'information', respectively. Information, then, can be understood as an act of producing the state of knowledge. Information is connected to the means of transmission, to the kinds of signals through which this transmission takes place. Computer science can be seen as providing technological ground for the application of information science (ibid., p.14), while information systems are central for processing data. Computer science

encompasses a manifold of research domains ranging from development, design, and application of algorithms and visualisations to software development, to human-computer interaction. Information systems research, in turn, while including information technologies—both hardware and software which provide technical foundation for information systems,—is also crucially concerned with the kinds of problems an information system is designed to solve and its embeddedness in the organisational processes (Laudon & Laudon, 2020, p. 18). From the perspective of information systems research, 'information' consists of formatted digital data that "is meaningful and useful to human beings" (ibid., p. 16). The data, then are digital documentations of events, actors, or artifacts "occurring in organizations or the physical environment" (Laudon & Laudon, 2020, p. 16). This semiotic perspective on information draws on works by Charles Sanders Peirce and Ferdinand de Saussure and allows to translate the concept of 'information' into technical means through information systems (Stamper, 1996). "Data is what is left of information when the cultural context is abstracted away. What is left then is the material aspect of information" (Nake, 2002, p. 48), thus creating the "machine view [distinct] from the collective and individual views" (ibid., p. 49). This, then, allows various manipulations with the data, based on their representation, structure, statistical properties, etc.

Using various algorithmic techniques, data that function as an input for computation can be processed according to certain computational rules to produce an output of such algorithmic analysis (see Heuer, 2020, p. 32). While algorithms follow certain computational rules—sets of steps required to transform input data into an output—other techniques, currently often subsumed under the umbrella term 'AI' draw on machine learning (ML) techniques. In contrast to algorithms as described above, ML-based algorithms do not follow certain predefined rules but are trained in an optimal way according to certain criteria (ibid. p.33). That makes ML-based systems challenging for analysis as the algorithms' designers do not have immediate control over the output such algorithms produce. Nevertheless, in our current, datafied societies, digital data can be seen as carriers of information, while algorithms are used to ascribe new meaning to data so that the resulting information can be put to use as applied knowledge. What this meaning is and how it is being ascribed, however, cannot always be rendered visible or even understandable. Discussing their experiences in Sciences Po médialab, Venturini and colleagues (2015) offer critique of the metaphor of 'data mining' as algorithmic data processing. For Venturini et al., this metaphor implies that in the course of such processing applied algorithms can distinguish clearly 'information' from 'noise' in the data (p.19). According to the authors, however, exactly this distinction is usually the result of such an analysis rather than its start. As I will show in the section 2.2 discussing further concepts and research domains interested in digital data, some of these domains (e.g. fairness, accountability, and transparency studies) are specifically interested in this kind of methodological analysis of the data and techniques for their processing, criticised by Venturini and colleagues.

Other scholars such as Boellstorff (2013), Kitchin (2014c), and Drucker (2011) critique the word 'data' itself, noting how the meaning of a Latin word *datum*, grammatically connected to the verb *dare*, is 'to give' or 'given'. The authors argue against viewing data as given and suggest using a 'more fitting' word 'capta' derived from Latin 'to take': "those units of data that have been selected and harvested from the sum of all potential data" (Kitchin, 2014c, p. 2). Against this background, the next crucial point in understanding data is, hence, grounded in the origins of the word: the common understanding of data is something 'given' which can be extracted from reality, whereas actors extracting these data according to their goals are only implied. From a more critical perspective, however, data are rather being constructed in order to produce certain representations. Bowker (2005) and later Gitelman (2013) made this argument explicit by defining data as not 'raw', but rather 'cooked', configured, 'imagined' in the specific field of application.

In social sciences, boyd and Crawford (2012, p. 663) published a seminal critical piece, in which they focus on data as an interplay of technology, analysis, and mythology. They particularly address data as 'big' in their account. Puschmann and Burgess (2014) discuss in detail differences between the terms 'data' and 'big data'. Summarised briefly, in contrast to the etymological origins

and historical use of the word 'data' discussed in this chapter, the term 'big data' originated in industry: "[b]ig data marked a suggested shift from relational database management systems to platforms that offered long-term performance advantages over traditional solutions." (Puschmann & Burgess, 2014, p. 1694) As Mayer-Schönberger and Cukier coined the term 'datafication', as we currently use it in social sciences, in 2013, their analytical attention was directed first and foremost to the so-called 'big data'.

> "The era of big data challenges the way we live and interact with the world. Most strikingly, society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing *why* but only *what*" (p.6-7, original emphasis).

Big data are about to change the world, "[h]arnessing vast quantities of data rather than a small portion, and privileging more data of less exactitude, opens the door to new ways of understanding" (Mayer-Schönberger & Cukier, 2013, p. 18). According to this, data are ontologically similar to natural and social phenomena and processes they are meant to represent. Epistemologically, then, data can be understood as 'facts' and 'truths' about these phenomena and processes. Analysing data with appropriate techniques of identifying patterns, then, methodologically allows to make assumptions about the world based on these data. Datafication, in this line of thought, is about "taking information about all things under the sun — including ones we never used to think of as information at all […] — and transforming it into a data format to make it quantified" (ibid., p.15). As Mayer-Schönberger and Cukier themselves predicted, 'big data' and 'datafication' became "a victim of Silicon Valley's notorious hype cycle" (p. 7), as both tech companies, media, and states adopted the two concepts for their practices and discourses. In this chapter, terms 'big data' and 'data' are used interchangeably, although preference is given to the term 'data'. The reason for such use of both terms lies in the blurring of borders between the two in academic knowledge production in the broad literature in datafication scholarship, media studies, and studies of algorithms and various automated systems.

Due to the commercial origins, the term 'big data' received critique across various academic domains (e.g. Crawford et al., 2014). Despite such critique, academics continue using the term even recently (e.g. Krasmann, 2020; Nikunen, 2021). In their seminal, same-titled article boyd and Crawford (2012) formulated "critical questions for big data" contributing greatly to the development of a critical discourse around digital data in social sciences. They argue that data do not 'speak for themselves' and require interpretation from their emergence in a specific information system and later in its commercial or scientific use:

> "[t]he design decisions that determine what will be measured also stem from interpretation. For example […] there is a [subjective] 'data cleaning' process: making decisions about what attributes and variables will be counted, and which will be ignored" (boyd & Crawford, 2012, p. 667).

Data are, then, not representations of 'facts', but are rather constructed within socio-technical systems and in accordance with the goals and values guiding involved actors. Borgman (2015) makes the same argument, conceptually developing it from the multiple definitions of data such as so-called 'Vs'—volume, variety, velocity, and later also veracity of data that allow speedy processing, reproduction, transfer, and use (p. 6). These dimensions cannot be seen as much as properties of data as they are properties of the socio-technical systems in which these data are being processed. If data are defined through the ways they are operated with, processed, and used, so Borgman argues, the core question is not about what data are, but when. The 'bigness' of data, too, relates to the ways data can be put to use or not; furthermore, it is at odds with manifold situations where data about certain people or processes are missing, incomplete, broken, or unavailable. Expanding on Borgman's question about 'when are data' in this way makes explicit the role of socio-technical systems in which these data are put to use, and the work required from various actors in order to use these. Such a critical understanding of digital data denies any notion of data's neutrality and highlights how data come to be only in the process of their use for specific (research)

goals. Crucially, data cannot be 'found' in the empirical field, but rather they are made through specific practices of their production and use (see van Atteveldt & Peng, 2018 for the secondary use of data). Various human and material actors fill data with meanings (e.g. Dourish & Gómez Cruz, 2018), producing information and knowledge as these data find their practical application in the manifold of societal domains and practices.

Other scholars like Pink et al. (2018) and Lupton (2018) further develop the idea of data as a social(-material) construct and consider embodiment and material aspects of data in their research. Essentially, data, therefore, are not sole representations of the material realty, but rather an interpretation thereof (e.g. Bowker, 2005; Gitelman, 2013). Data can be understood as a 'social product',

> "selected according to diverse social factors […]; organizations and institutions within which data are collected […]; political, financial and marketplace contexts […]; the material availability of devices to enact the generation of data; and the legal and ethical frameworks, technical standards, laws and regulations that govern all stages of the production and use of data"(Williamson, 2017, pp. 29–30).

Viewing data as social product opens perspectives on the multiple challenges of the enthusiastic definition of datafication processes, allegedly allowing to better understand social processes. The 'bigness' of data, their availability, their use, and usefulness now can be put in the context of other social processes, also including discrimination, inequalities, and injustices that also shape how data come to be and what roles do they play in societies.

Critical scholars not only argue against understanding data as something given and factual, but also as objective, and truthful, expanding on the long-standing research traditions like sociology of numbers, research on quantification or metrics (e.g. Christin, 2020; Gorur, 2018; Lippert & Verran, 2018; Mau, 2017; Passoth & Wehner, 2013). Such studies address various analogous and digital data epistemologically, challenging their presumable objectivity. Lippert and Verran (2018) explicitly bring together number, algorithm, and data studies switching attention from the data and numbers themselves to the practices of knowledge production—methods and techniques of analysis, sometimes automated or algorithmic[1]—that make data appear objective and truthful. This argument echoes the point that Puschmann and Burgess (2014) make in their analysis of metaphors of big data as imaginaries of data produced in big tech industry. Being aware of such imaginaries and the ways they enter academic knowledge production is crucial for a critical understanding of data and datafication processes. For example, Mau (2017), makes a point in demonstrating how numbers and metrics are de-contextualising phenomena and processes they mean to count (p. 27). He exemplifies the de-contextualising power of numbers by elaborating on rankings. According to Mau, public representations through ranking may be fetishised[2], and instead of "being good" the efforts are put into producing accounts that are "looking good".

Having discussed various ways of understanding data, I finally turn to the manifold typologies of data used in social sciences and how these data originate. For example, there are networked data—generated in interactions of people with socio-technical systems, for example in form of software interfaces (e.g. Airoldi, 2021; Perrotta & Williamson, 2018 using such kind of data). Prevailing attention to such transactional data, however, has been criticised by some scholars like Beer and Burrows (2013), who propose to look beyond these, distinguishing digital data and their archives according to the content like everyday activities, opinions and viewpoints, or crowdsourced knowledge. Kitchin (2014c, p. 4) distinguishes between different forms and types of data (qualitative, quantitative & index, attribute, metadata), their source (e.g. captured, derived, exhaust, and transient data), structure, and producers of data. Zuboff (2015) additionally identifies

[1] For a critical discussion of algorithms and artificial intelligence (AI) as analytical techniques of data science see for example Heuer et al. (2021).

[2] For a discussion on data as fetish, see e.g. Chun (2011).

economic transactional data, sensory data, corporate and government data, and surveillance data (e.g. from public surveillance cameras to Google Street View) (p. 78-79). In contrast to the former types of data, the latter ones discussed by Zuboff already include contextualisations in regard to the organisational entities owning these data or to the technologies through which these data are being generated. In computer science, other categorisations are relevant, for example in relation to how these data are structured, whether these data are homogenous or heterogenous or the types of data values (characters, timestamps; binary or string data, etc.). Latter is specifically important for storing such data digitally, as it gives an account on the amount of Bytes reserved for the data of each type. Finally, additional attention also require metadata—"data about data" (Borgman, 2015, p. 66).

The many kinds of data discussed here do not only provide different conceptualisations of what digital data are and do, but also direct attention to various ways in which digital data originate. They originate in different socio-technical systems such as technological, material infrastructures including sensory devices (e.g. embedded into self-driving cars or smart cities' infrastructures), commercial organisations or public institutions and their information systems, or corporate and state surveillance networks. For example, the metadata mentioned above are usually assigned to certain digital artifacts automatically: a digital photograph may contain metadata such as a time stamp when the photograph was taken, the type of the digital camera, and characteristics of a camera lens. Other digital devices such as devices used in various domains of natural, environmental, medical sciences, or city planning include different kinds of sensory systems that also allow data generation. For example, Rajão and colleagues (Rajão & Jarke, 2018; Rajão & Vurdubakis, 2013) draw attention to the deforestation in Brazil by studying various kinds of data such as satellite images or data from GPS devices. Another example of sensory data currently increasingly discussed and critiqued by critical scholars, particularly in the domain of education technology research, concerns so-called neurotechnology that aims at producing data from brain activity to be analysed according to certain capacities of the brain (e.g. Williamson, 2019). In health and medical research, however, such and other technologies used to 'datafy' human body are not only widely used, but also essential. For example, Lindén (2021) in her study of patients' groups discusses how these groups campaign for more research (and resulting data) on gynaecological cancer, while multiple publications in the special issue by Coopmans and McNamara (2020) discuss explicitly or implicitly how health data are used and sometimes misused to care for patients.

Other information systems specifically generate data based on human activities within these systems which subsequently can be used for further analysis. For example, some scholars study log files—a technical procedure historically used primarily for identification of system errors that can document which kinds of actions have taken place in a given information system, e.g. based on this system's user activity (e.g. Schulz & Breiter, 2012). The term digital or data 'traces', in contrast, foregrounds the networked, digital data that people leave as footprints of their media use (Hepp et al., 2018, p. 439). Some scholars like Thylstrup (2019) criticise the lack of definitions of digital traces and point out sometimes colloquial use of the term:

> "American sociologist Matthew J Salganik (2017: 71), for instance, treats digital traces as the by-product of people's everyday actions, simply noting that 'I've used the term of digital traces, which I think is relatively neutral' compared with similar terms such as digital footprints and digital fingerprints. French sociologist Franck Cochoy et al. (2017: 4) use the notion in a similarly loose fashion [....]" (p.3).

Despite the diffuse definitions, many authors have incorporated the concept of digital or data traces in their empirical research, e.g. to analyse practices of self-tracking (e.g. West, 2019) or social media and platforms[3] (e.g. Alexander et al., 2018; Grenz & Kirschner, 2018). As Hepp and colleagues

[3] In my thesis, I use the term 'platforms' either in relation to social media sites or when directly referring to the work of other scholars applying the term. For critical interrogation of the term 'platform' see for example Gillespie (2010).

(2018) further explain, digital traces sometimes are left consciously, for example by uploading content in a cloud, and sometimes involuntary, for example by browsing in the web (p.239-440). Other authors, e.g. Venturini and Latour (2010) argue that by observing and analysing digital/data traces, social processes can be observed:

> "A criticism that has often been addressed at the analysis of digital traces is that these traces regard a non-representative sample of society. From the point of view of the traditional social sciences, this is certainly true. It is well known that digital literacy is not uniformly diffused in society (men, young people, and those with high levels of education are generally overrepresented in online samples). Still, this disproportion is a problem only as long as we insist on treating digital data as if they were survey data. The advantage of the new methods is that they allow tracing the assemblage of collective phenomena instead of obtaining them through statistical aggregation" (n.p.).

Similarly, in their methodological contribution on digital traces, Welser et al. (2010) follow a structural approach for computational social science and attend primarily to the 'structures of action' rather than to its content. Such an understanding of the role of digital traces, however, does not take into account fully their social origins and their interpretive, relational, rather than representational character. In this hindsight, questions of users'—persons who generate digital traces with their behaviour—agency are being raised. These conceptualisations, although arguable, render visible how digital data and traces decontextualise individuals and their behaviour, for example if actions are being analysed without attention to their content. Digital data become separated from the situations in which they have been produced and require both interpretation and renewed 'locating' (Haraway, 1988; Suchman, 2002) in these situations. I proceed with a brief overview over multiple understandings of digital data dominant in different academic discourses: I put digital data in context of academic discourses in social sciences. An interplay between digital data, datafication processes, and related socio-technical systems is a recurring theme in my thesis.

Critical research on digital data addressed so far illustrates a different epistemology than that articulated by Mayer-Schönberger and Cukier (2013) in their book. The introduction to critical scholarship presented here is neither extensive nor full—and will be extended in the following, however its purpose in this chapter is in addressing digital data not only as an empirical phenomenon brought about through implementation of various socio-technical systems, but also as an element of knowledge production, not at least that of academic knowledge. In practices of academic knowledge production, digital data become intertwined with new, research-related socio-technical systems, as these data are being incorporated in research practices. As I will show in the next sections of this chapter, for research on datafication, it poses a challenge: digital data are not only considered as parts of empirical datafication processes, but also are an essential element of research processes. It is not at least due to this challenge that I continue speaking about 'data' rather than information in my thesis. Before turning to these methodological and practical issues of datafication research, I extend on the critical approaches to understanding digital data, presented in this section, and introduce further terms and concepts relevant for research on datafication. In the next section, I set out to explore how, within different academic discourses, data make a difference in the world, what difference do they make, and how the manifold of concepts of 'datafication' address these differences.

2.2 How do digital data change the way we understand the world?

In the first section of this chapter 2, I provided a brief overview of different ways to understand the term 'data' from the perspectives of such academic disciplines as information science and social sciences. This section of the chapter continues this discussion by introducing various concepts related to data widely adopted in social sciences for dealing with and countering their elusiveness (Borgman, 2019). These concepts highlight how digital data change the way we understand the

world, putting data into relation to different aspects of social lives such as economy, surveillance, identity, and (mis)representation.

So, digital traces are not simply left by people; the kinds of traces people leave in information systems are defined through their design. After being 'left' these traces are being processed according to the categories of these formerly defined traces and combined in use profiles, for example by social media companies. By combining various traces of the same user, new data about their online behaviour are being produced, further processed, and sold, which is sometimes also addressed as 'data production'. This notion of 'data production' serves to distinguish between the practices of technology providers, processing the data ('production') and practices of users (who also generate data, e.g. digital traces, with their behaviour, and user-generated content) and the effects on the users. Following Borgman's (2015) introductory argument in an analysis of digital data, "[t]heir value lies in their use" (p.3), commercial companies such as social media and other technology providers build their business models on value extraction from digital data. In academic discourses, this has been addressed from a variety of perspectives, highlighted through notions like 'surveillance capitalism' (Zuboff, 2015), data as capital (e.g. Sadowski, 2019; West, 2019), and an ideology of 'dataism' applicable not only to the commercial corporations, but also to academica and states (van Dijck, 2014). For example, Zuboff (2015) describes surveillance capitalism as a "deeply intentional and highly consequential new logic of accumulation" that "aims to predict and modify human behaviour as a means to produce revenue and market control" (both p. 75). In the core of such accumulation is the ability of companies—manifested in their financial, personal, methodological, and knowledge resources—to combine multiple data points in order to extract patterns that allow further economic transactions and, ultimately, multiplication of the companies' revenues.

Some researchers address profiles built from manifold combinations and linkages of various data points about an individual as 'data doubles' (Jones & McCoy, 2019; Lupton, 2012; Ruckenstein, 2014). As Vallee (2020) elaborates,

> "[t]hese 'doubles,' mutable as they are, return to us in the form of partially recognizable versions of ourselves. The most literal example of a data double would be an avatar, or a YouTube recommendation, or even broader composites that contribute to the content of television shows based on audience tastes determined from aggregated viewing habits" (p.3).

The quote illustrates how certain representations of an individual's behaviour are addressed as data doubles. Cheney-Lippold (2017), in turn, discusses the "measurable type"—the kinds of data available that allow coming to certain conclusions about people or things these data are meant to represent; the process of coming to these 'conclusions' and giving meaning to data is addressed as "profilization" (p. 87). Cheney-Lippold underscores the notion of a measurable type as a kind of classification and a model, putting forth the argument that an interpretation of such a measurable type can be understood in terms of surveillance and fitting subjects (of datafication) into the frames of profiles set for them. In academic research, discussions about data double are related to further concepts: 1) surveillance or so-called dataveillance (Lupton & Williamson, 2017; van Dijck, 2014) and 2) the role of the subject and the relations between data doubles and identities of actual people being datafied (Jones & McCoy, 2019). Here, I only briefly touch upon the key points of surveillance and digital identity studies as both are academic fields and topics of their own. My aim here is to highlight further essential aspects of digital data in relation to these. For this purpose, the relation between data and surveillance can be addressed through the notion of "[d]ataveilalnce—the monitoring of citizens on the basis of their online data" (van Dijck, 2014, p. 205). According to van Dijck (2014), in contrast to surveillance, dataveillance has no specific purpose, rather the purpose may change over time and be developed out of the obtained digital data. Through dataveillance, datafied representations of individuals and groups are used for value extraction, raising not only concerns about privacy (widely discussed across multiple academic fields, see e.g. Lai & Flensburg, 2020a; West, 2019), but also about visibility (e.g. Flyverbom et al., 2016; Neumayer et al., 2021),

agency (e.g. Kennedy et al., 2015), transparency, and accountability (e.g. Wieringa, 2020). Similar issues are also being tackled by many other scholars who provide us with extensive analyses of the interplay between digital data and individuals' identities, addressing questions like how people become data, what is the relation between the selves and their data (e.g. Forlano, 2016, 2017; Koopman, 2019; Lupton, 2012; Papacharissi, 2018). These questions and concepts vividly illustrate recursivity of digital data in relation to people and their selfhoods. As Koopman (2019) summarises in the introductory chapter of his book,

> "our data are not mere externalia attached to us from which we might detach our truer selves as we please, but are rather constitutive parts of who we can be. Who we are is therefore deeply interactive with data. We are cyborgs who extend into our data" (p.8).

As representations of certain aspects of individual identities and behaviours, data are being acted upon. Lupton (2012) explicates this point by stating that "[t]he flow of information, therefore, is not one-way or static: it is part of a continual loop of the production of health-related data and response to these data." (p. 237) This recursive loop reminds us that digital data are not only *of* the social (as social product, defined above), but also are *being* social, able to reconfigure social lives.

Current critical scholarship demonstrates, however, that attending solely to data rather than to actual people that are meant to be represented though them may lead to misrepresentation, silencing, and discrimination—outcomes contrary to those of 'representation' (Whitman, 2020, p. 3). Multiple, particularly feminist, queer, and scholars of colour have shown that often underrepresented, marginalised, and 'othered' groups of people are not represented in the data upon which political and governance decision-making takes place (e.g. Benjamin, 2019; D'Ignazio & Klein, 2020; Draude et al., 2019; Noble, 2018; Perez, 2019). These scholars not only shed light on the biases and inequalities reiterated through digital data and algorithms for their processing, but also on different kinds of data value than economic value extraction. They provide examples of absent and missing datasets that would have contributed to a better representation of certain groups of people, for example women (for other research also concerned with different kinds of data values see also Bolin, 2022; Fiore-Gartland & Neff, 2015). Discussing such absences and othering in terms of knowledge production, Leonelli and colleagues (2017) elaborate on

> "an ambiguity and a strategic relationality to shadowing processes that parallels the relational nature of data and the multiplicity of motives, goals, and conditions through which data may be construed as (in)significant, partial or complete, (un)intelligible, or (in)accessible" (p. 194).

This argument makes a loop back to the introductory definitions of data given in this section, further putting forward the politics of data, for example enacted in the kinds of data accounted for or not. Some scholars addressing digital data research from a feminist perspective, also turn to the notion and concept of 'care': they investigate how various human actors relate themselves to digital data not merely by unconsciously leaving traces of their activities in information systems, but also relate to data affectively and ethically (e.g. Baker & Karasti, 2018; Fotopoulou, 2019; Zakharova & Jarke, 2022). Taking such a position, these authors attempt to provide an analysis of data as an inherent part of our societies, closely intertwined with people and their lived social realities, rather than as something abstract and external to these realities.

At the intersection of social sciences more broadly and computer science, research fields have developed which specifically address the challenges of bias in data. For example, fairness, accountability, and transparency (FAT) studies particularly focus on bias in various automated systems. According to Peña Gangadharan and Niklas (2019, p. 884), initially, FAT studies were concerned with bias in automated systems and the ways to reduce it. The authors argue, however, that "fairness, accountability, and transparency studies fail to air their value-based assumptions about antidiscrimination or fairness, remaining inexplicit about their allegiances to any one normative framework." (ibid., with reference to Binns, 2018). Rather, so Peña Gangadharan & Niklas (2019), datafication and algorithmic processes considered in FAT studies can be seen as technological processes that may be 'fixed' with a technological solution such as de-biasing.

Further, as I have shown in the previous section of this chapter, discussing the distinctions between ML-based (also 'AI') and other algorithms, algorithmic transparency is not achievable for the former as even the designers/researchers themselves cannot account fully for what kinds of output these systems would produce and why.

A concept of 'data assemblages' places digital data within the manifold aspects of socio-technical systems such as economic, political, social, and technological (Kitchin, 2014c; Kitchin & Lauriault, 2014). Widely applied across various domains of social sciences and in critical data studies, data assemblages as an analytical concept provide an answer to the question of how to address digital data critically, connecting them to the society at large (Iliadis & Russo, 2016). Other scholars also offer critique of data assemblages, addressing the lack of attention to human agency or relations between the elements within a data assemblage (Heeks & Renken, 2018, p. 97).

Similar to the application of the concept of data assemblages, attending to data infrastructures allows to put digital data in relation to the material, social, political, and economic processes, as Gray and colleagues (2018) elaborate in the following:

> "[l]etting go of the notion that data does nothing more than show us how things are, we can attend to the social, historical, cultural and political settings in which it is created and used and which framings such infrastructures introduce to the data. To this end, Bowker and Star (1999: 34) call for "infrastructural inversion": bringing the background work involved in the making of data into the foreground and hence we can study the social practices which databases both reflect and enable, such as quantification, classification, commensuration and calculation" (p.3).

Gray et al. (2018) build here on the research on infrastructures and classification conducted by Jeffrey Bowker, Susan Leigh Star and Karen Ruhleder (e.g. Bowker & Star, 1999; Star & Ruhleder, 1994, 1996), who understand infrastructure as "a working relation", occurring "when local practices are afforded by a larger-scale technology, which can then be used in a natural, ready-to-hand fashion" (Star & Ruhleder, 1996, p. 114). In relation to data and datafication processes, various authors have adopted an infrastructural approach to analyse "datawork means of which social media platforms standardize and compute user participation" (Alaimo & Kallinikos, 2017, p. 176), data movement within and across data infrastructures (e.g. Gerlitz et al., 2019; Weltevrede & Jansen, 2019), or frictions occurring as data move and are being moved (e.g. Aula, 2019). As the examples illustrate, the notion of data infrastructures foregrounds movement of data and the work 'behind the scenes' required for that movement[4]. This work encompasses practices of standardisation and categorisation (see e.g. Piattoeva & Saari, 2020 for an analysis in context of academic knowledge production), and renders various data visible or invisible (e.g. Flyverbom & Murray, 2018). For the categorisation work, a specific kind of data—metadata—are central, as they define how other data can be clustered for further processing and use, set standards, and allow for interoperable communication between various systems (e.g. Kubicek et al., 2019). The work of categorising data and setting them on the move (Jarke & Zakharova, forthcoming) is political (Bates, 2018). Ultimately, data politics concern questions of power (who decides what data are important and why, who has access to these data, who makes which decisions upon these data), of accountability (who counts and who is counted, who is being held to account according to certain data), knowledge production (who makes which data available for further re-use), and elicit new forms of power relations between various actors creating new subject and object positions (Prietl, 2019; Ruppert et al., 2017). As most recent literature illustrates, even critical scholarship can be "compromised" and not able to withstand the influence of some actors who define what and when data are according to their, often commercial, interests (Whittaker, 2021, p. 52).

[4] Besides the concept of work, other approaches can be used to describe what is happening 'behind the scenes' of datafication processes, such as e.g. 'doing' data. See for example Wanka and Gallistl (2018) for conceptual understanding of 'doing'.

Related work presented here should provide an answer to the question in the title of this section: how do digital data change the way we see the world? This answer, however, is not a simple one. Outlining various concepts of digital data and their relations to algorithms and data infrastructures this section briefly reviews the manifold of answers to this question. As Ruppert and colleagues (2017) notice,

> "[n]otably, attention has started shift to a focus on computation and analytics such as algorithms, machine learning, and artificial intelligence. Yet, data remains a key matter of concern as both the product and condition of computation and analytics" (p. 2).

There is, however, no one definition of digital data. Rather, data acquire meaning in the process of their use in relation to ontological, epistemological, and methodological backgrounds of the actors operating with data, empirical fields in which they operate, and their goals. Data can be understood as assemblages that acquire their meaning in socio-technical systems within and across which these data are being put to use. These socio-technical systems include not only various technologies such as software or data infrastructures, but also (automated) methods of recombination and analysis applied to these data such as algorithms, and broader economic, political, and social contexts. Data are, therefore, not representations but interpretations of reality, nor are they neutral or objective; instead, data are relational and of the social. With that, data can reproduce biases and inequalities, include some people and exclude others. With that, the title question itself becomes controversial, as the 'we' changes along with the ways different people are datafied or not. Data are recursive as they are not simply extracted and combined into datafied profiles of individuals, organisations, and whole societies, but also affect these in a manifold of ways as they are being acted upon individually and collectively. Some data, such as metadata, also categorise and set standards for other data. As long scholarly traditions in social sciences and STS demonstrate, data and categorisations they produce about other kinds of data can be othered, missing, and broken instead of being big, truthful, objective, or open and available for re-use (e.g. Borgman, 2015; Bowker & Star, 1999). Data are, therefore, political, embedded in various political discourses and, thus, have their own politics when put to use by different actors. The main argument of this section of my thesis is that these data politics are also a part of academic knowledge production—the ways researchers see[5], conceptualise, and ultimately describe the world. The manifold of ways in that data are intermeshing with the existing views, concepts, and descriptions illuminates how data and datafication processes complicate knowledge production in research and practice. I discuss this latter point in more detail in chapter 3. Altogether, the workings of the data and the ways of doing data in socio-technical systems can be broadly understood as datafication processes. Expanding on the notions of data reviewed here, in the next section, I elaborate in more detail on various understandings of the concept of 'datafication' and further relate it to other academic discourses.

2.3 What is datafication?

So far, I have shown the manifold of concepts related and constitutive to understanding of datafication processes (e.g. previously big data, then also digital traces, quantification) and the variety of concepts of datafication highlighting various aspects, elements, and implications of datafication processes. Several years after the concept of datafication has been coined, the datafication scholars across various domains of social sciences and other disciplines embarked on a journey to produce more situated definitions and weave together the technological and sociocultural processes enabling datafication. In this section, I provide an overview over different perspectives on datafication processes and their relations among each other and to further, related concepts.

[5] For the critique of the metaphors of 'vision' see for example Barad (2007), Strathern (2000), Star & Strauss (1999), and Garforth (2012).

I began this chapter by recounting an argument by Mayer-Schönberger and Cukier (2013) about the revolutionary role of big data in understanding the world: if only we had data about everything in the world, we could understand how it works. As my brief review of critical data scholarship in the previous section shows, it is neither possible to datafy everything and "data deserts continue to exist" (Kitchin, 2014c, p. 150), nor does datafication, *per se*, further our understanding of the societal processes. The emphasis on 'understanding' of the world in the early, empiricist (Kitchin, 2014c, p. 130) definitions of datafication, however, requires more attention. Critical scholars address datafication as an episteme (e.g. Beer, 2016) or an ideology (e.g. van Dijck, 2014). For example, Kitchin (2014c) and Beer (2016) refer to 'big data' as a concept or a rationale used by the industry to promote data analytics (p. 6). Others explore metaphors used in discourses about digital data and datafication in order to reveal the kinds of knowledges these data are meant to provide scholars studying them (Pangrazio, 2020; Puschmann & Burgess, 2014; Stark & Hoffmann, 2019), drawing attention to the abundance of 'natural' metaphors like oil and economic metaphors like asset or utility which reinforce the empiricist, positivist epistemology. As Mejias and Couldry (2019) argue, "much-used metaphors that equate datafication to other extractive processes help to further obscure, not uncover, these power relations" (p.4).

Rather than using such metaphors, many scholars have discussed datafication as a continuation of different practices that existed prior to the intensive rendering of various aspects of the world into digital data that we observe today (see Koenen et al., 2021; Ruppert & Scheel, 2021). In a historical overview over the roots and development of datafication processes, Koenen and colleagues (2021) discuss record keeping beginning in ancient Greece and briefly follow it through centuries and European regions as an example of pertaining data practices (p. 140-141). With this and other examples such as discussing the role of punch cards for automated data processing in Nazi Germany, the authors illustrate how "the production, collection, and processing of data not only predate digitalization but also, in the immediate decades before the digital revolution, produced exclusive infrastructures, knowledge orders, and practices" (ibid., p. 152). Other authors also point to the role of data in national statistics such as censuses (e.g. Didier, 2022). In a similar vein, Ruppert & Scheel (2021, p. 292) in their analysis of statistical data practices of making European people also call to question categories of population statistics and academy, meaning also how they came to be. What these examples illustrate is how the methods of quantification usually serve certain purposes and are performative to the concepts about the society developed with the help of these numbers. Similarly, datafication processes are performative to the societal domains that are being datafied. Such performativity of the concepts and methods to the academic and applied knowledge about our datafied societies is the main recurring theme of my thesis. While I address the related methodological questions beginning with the section 2.4 of this chapter, the aim of this section is to illustrate though a review of related work how various conceptual definitions of datafication draw analytical attention to different aspects that require certain empirical and methodological choices from the scholars interested in studying datafication.

Currently, scholars interested in datafication processes work across many domains and fields of social sciences. For example, in regard to what societal practices and processes are being addressed in datafication scholarship, table 2-1 provides an overview over current research, ranging from studies of the everyday, health and self-quantification, to the studies of data activism, education, welfare states, and justice. This is not a comprehensive list as the societal domains and practices affected by datafication processes are countless. Most of research listed here can be described as critical according to a comprehensive literature review of datafication scholarship by Flensburg and Lomborg (2021). Such critical scholarship relies on a relational understanding of data, while datafication can be understood processually, as an enactment in which different actors translate various societal processes into data and back according to their goals and in relation to the contexts and situations in which this enactment takes place. For example, such a processual view on datafication is elaborated by van Dijck (2014, 2017). According to her, through the process of

datafication, various societal domains increasingly become defined through quantitative data. Making a similar argument, Cheney-Lippold (2017) argues that datafication can be seen as connecting various data points with power (p.107) through the notion of a 'measurable type'. With particular focus on identities, the author argues that the ability to create 'profiles' for these measurable types is driving the big tech industry's ability to put data to use and give them 'meaning'. For Cheney-Lippold, this process of giving data 'meaning' is not only an element of big tech economy, but also a question of changing subject and object positions of individuals and their profiled identities. Mejias and Couldry (2019) further emphasise economic value extraction and commercial purposes that are guiding an enactment of datafication processes For them,"[t]he term "datafication" implies that something is made into data. What that something is, and what the processing comprises, are matters that need to be put into context" (p.1). The understandings of digital data discussed in the previous sections, however, also highlight that digital data also need to be understood in the context of their use/application.

Table 2-1 List of research domains and societal practices addressed in current datafication scholarship.

Audience research	(e.g. Livingstone, 2019; Mathieu & Jorge, 2020; Mathieu & Pruulmann Vengerfeldt, 2020)
Data activism, Global South perspectives, and migration	(e.g. Lehtiniemi & Haapoja, 2020; Mattoni, 2018; Milan & Treré, 2019; Witteborn, 2021)
Data literacy	(e.g. Markham, 2019)
Data journalism	(e.g. Dowd, 2016; Steensen, 2019)
Education on different levels from early years education to schools to higher education	(e.g. Jarke & Breiter, 2019; Lewis & Holloway, 2019; Williamson, 2017)
Environment and climate, smart cities	(e.g. Bates et al., 2016; Kitchin, 2014a; Madsen, 2018; Rajão & Jarke, 2018);
Everyday practices	(e.g. Karlsson, 2019; Kennedy et al., 2020)
Feelings, affect	(e.g. Kennedy & Hill, 2018; Lupton, 2018; Sellar, 2015)
Governance, public administration, welfare state, and algorithmic decision-making	(e.g. Dencik & Kaun, 2020; Kaun, 2021; Velkova & Kaun, 2019)
Health and self-quantification	(e.g. Lupton, 2016; Schüll, 2016)
Justice	(e.g. L. Taylor, 2017)
Organisational practices and data-driven decision-making for organising	(e.g. Alaimo & Kallinikos, 2021; Leonardi & Treem, 2020)
Social media, platforms, and apps	(e.g. Bucher, 2018; Helmond, 2015; Lai & Flensburg, 2020b; Pybus & Coté, 2021; Weltevrede & Jansen, 2019)
Work and labour	(e.g. Sánchez-Monedero & Dencik, 2019)

Together with Hepp, Couldry (e.g. Couldry & Hepp, 2017; Hepp, 2020) defines datafication as an aspect of 'deep mediatisation'. Deep mediatisation, here, stands for the latest wave of mediatisation—a process reflecting the interrelations and dialectic transformations of media and society, interrelated with each other. The waves of mediatisation describe historical transformations of media, while deep mediatisation reflects the current addition of digital data and related datafication processes to the media landscape. 'Deep mediatisation' reflects the acceleration of such transformations and the relevance of media to the societies and cultures. Datafication can be understood, then, as an aspect of deep mediatisation,

> "offering the chances for new, quantified forms of 'reflexivity' through the ongoing data that individuals produce, but it is an aspect deeply driven by the needs of media and data

> industries themselves, and brings costs quite different from those previously associated with reflexivity" (Couldry & Hepp, 2017, p. 219).

In a later take on the concept of deep mediatisation and datafication, Hepp (2020) derives definitions of both concepts from the history of different media, highlighting various technological aspects of these such as the materiality and institutionalisation. As the definition provided here suggests, datafication here is tightly intertwined with technological transformations and possibilities of data processing. Similarly, in the research domain of education technology Jarke and Breiter (2019) define datafication as a process and means of measuring and defining social life in numbers. Williamson (2017) proposes a definition focusing specifically on the constructed character of digital data and the manifold of social, political, and technological elements involved in such construction. Particularly within education research or research on datafied childhoods, analytical attention to the negative implications of datafication processes such as biased profiling and related future exclusions or privacy risks is strong (e.g. Bradbury & Roberts-Holmes, 2017; Mascheroni & Siibak, 2021). Studies of data activism also highlight the high stakes of datafication for individual and collective activist agency and resistance against exclusive datafication processes (Mattoni, 2018; Milan, 2019; Stephansen & Treré, 2019).

Prietl (2019) argues that attending to datafication processes from the perspective of bias or 'mistakes' in data happening during automated analysis creates a perception that digital technologies could be ultimately neutral if these biases and mistakes were to be eliminated (p.2). Definitions of datafication through technological developments and innovation provided above, thus, induce similar perspectives on societal implications of technology implementation and use. The concept and the emerging body of work on data justice counter this narrative of the achievable neutrality of technology. The concept of data justice rather expands on "frameworks for wellbeing, to post-structural influenced theories of justice, and to constructivist understandings of technology (as always value-laden)" (Peña Gangadharan & Niklas, 2019, p. 882). For example, Taylor (2017) reviewing different understandings of data justice illustrates that besides the negative aspects of relations between digital data and governance, "data technologies can provide greater distributive justice through making the poor visible" (p.6). As multiple conceptualisations of datafication provided in this section suggest, critical scholarship is to a great extent focused on the possible negative effects and harms of datafication processes. Departing from that, in her analysis of data justice frameworks, Taylor (2017) argues for reconciliation of perspectives highlighting negative and positive implications of datafication processes as being required to further not only academic discussions on datafication, but also their practical implementation and regulation. Attending to the societal aspects of data justice (e.g. Dencik et al., 2019; Heeks & Renken, 2018; L. Taylor, 2017), this body of work is also concerned with data governance (e.g. Abraham et al., 2019; Leonelli, 2019; Micheli et al., 2020), shedding light on the politics of data and the ways these are inscribed in data and automated systems. Alongside with the concept of data justice, many scholars address data colonialism (e.g. Birhane, 2020; Couldry & Mejias, 2019; Isin & Ruppert, 2019; Madianou, 2019; Ricaurte, 2019) and turn to the datafication processes in the Global South, instead of the already well-researched and dominating academic discourses about datafication in the northern hemisphere (e.g. Arora, 2016; Cinnamon, 2020; Milan & Treré, 2019). Yet other scholars specifically elaborate on data in the lives of People of Colour and Indigenous peoples (e.g. Mcglotten, 2016), providing important critique to the established conceptualisations of datafication processes and putting forward concepts such as data justice not only as an analytical framework, but also as an agenda for future research and practice. Some of the authors specifically focus on practices of 'resistance' against datafication processes (e.g. Ricaurte, 2019) and develop calls for action both for academic and civic actors.

Feminist critical data studies build on different feminist research traditions and aim at pinpointing when, where, and who is getting excluded or discriminated through datafication processes (Cifor et al., 2019; D'Ignazio & Klein, 2020; Gardner & Kember, 2021; Posner & Klein,

2017; Thompson, 2020). Feminist data studies are also committed to showing care and "taking sides, participating, acting, making a choice, taking a position, but without taking for granted a general or fundamental principle on which these actions would safely and coherently be grounded" (López-Gómez, 2019, p. 10). As the quote illustrates, feminist data studies—as feminist research more broadly—address datafication not solely as an episteme or an empirical phenomenon requiring analytical attention; rather, within feminist data studies new manifestos and calls for action are being developed that aim to address and counter the invisibilities, inequalities, biases, and discrimination that critical datafication scholarship has been able to point out so far. Other scholars also use concepts such as (in)visibility, surveillance, and transparency to address changing subject and object positions of people facing datafication processes (e.g. Couldry & Yu, 2018; Neumayer et al., 2021). A few authors turn to the role of (in)visibility of data and some aspects of datafication processes specific to academic knowledge production. For example, Leonelli and colleagues (2017) discuss data shadows (p. 194). Pink and colleagues (2018) explore 'broken' data, Blackman (2019) elaborates on the encounters with data 'ghosts', and Kjær et al. (2021) switch attention to absent data. Besides shedding light on who is 'othered' and mis-/underrepresented in the data, exploring datafication processes in terms of visibility and 'ghostly' encounters leads to methodological challenges for research and academic knowledge production. Here, distinctions between data as empirical phenomena and as research material in empirical studies of datafication are blurring. For example, Kjær and colleagues (2021) reflect in their analysis of Twitter data, how to conduct research knowing that some data cannot be gathered and will be missing from the dataset and argue for developing sensitivities to

> "the otherness at stake in the material under scrutiny: that there is no God's eye perspective, only multiple, partial, contradictory perspectives, some of which belong to the ghosts themselves, the voices that were not "captured"" (p.17).

This and other methodological questions that arise in datafication scholarship are addressed in more detail in the next section of this chapter.

Some scholars are rather interested in datafication as transforming "subjects, objects, and practices into digital data" (Southerton, 2020, p. 1), shedding light on the role of an individual in a datafied society. For example, Kennedy et al. (2015), Livingstone (2019), Dencik (2020) argue that rather than solely focusing on the oppressive aspects of datafication processes and research that renders individual people helpless, ignorant, and uninformed, more analytical attention should be paid to individual agency in datafied societies. Moreover, one could add, also attention to the productive aspects of datafication processes, how trivial these might sometimes be, also requires attention and explication. Kennedy and Moss (2015) further this discussion by addressing datafied knowledge production in relation to publics (typically described as 'users'). Milan and Beraldo (2019) elaborate on bottom-up politics of data that allow publics and users to exercise their agency in relation to data power. Tsai and colleagues (2020) discuss agency and challenges to it for some actors in context of learning analytics in higher education. The authors, many of whom develop their understandings of data power and agency within the field of data activism research, invite to think of other forms of agency apart from algorithmic oppression, e.g. as reflexive or habitual practice of laypeople as users. Interestingly, the term 'users' itself can be seen as an example of an analytical term imported into social research from other domains such as e.g. human-computer interaction; the application and meaning of the term in various fields of social sciences, however, is constructed differently. These scholars interested in individual actors and their relations to datafication processes observe that current research is often focused on the big actors accumulating data and power (such as technology providers), although the more pressing issue is to understand agency and power of other actors such as individuals/laypeople (see also Pruulmann-Vengerfeldt & Hörste, 2020). Some authors go beyond this argument proposing to attend to the affective implications of datafication processes and focus on feelings these evoke (e.g. Kennedy & Hill, 2018; Saifer & Dacin, 2021; Smith, 2018). Their research interest lies specifically in an investigation of

everyday activities of people facing and enacting datafication processes, for example through self-quantification (e.g. Karlsson, 2019; Schüll, 2016), encounters with and perceptions of data (e.g. Kennedy et al., 2020, 2021).

Some of the perspectives on datafication discussed so far focused primarily on individual actors, their perceptions and practices in light of datafication processes, while others—e.g. those working on the questions of data justice—explore datafication processes in relation to communities and groups of actors. Other researchers across various domains of social sciences and related domains of organisation studies or information systems research turn their attention to the analysis of datafication processes in relation to organisational practices (e.g. Alaimo & Kallinikos, 2021; Leonardi & Treem, 2020; Madsen, 2018). In this strain of research, datafication processes are being addressed in "production contexts" (Madsen, 2018, p. 10). Building on the set of critical questions to big data, formulated by boyd and Crawford (2012) and their attention to the processes of data processing and cleaning executed by technology providers, research on datafication and organisations explores organisational values, contexts, and practices and their role in 'organising' contemporary work and labour. For example, Alaimo and Kalinikos (2021) conclude their analysis of platforms as organisations with an argument about the co-constitution of organisations and technology:

> "it is no longer fruitful to treat technology as an exogenous force, separate from the organizational operations into which it is embedded. Organizations and technology co-constitute one another and have accordingly to be studied in tandem" (p.1387).

Education research with the manifold of recently published studies on datafication of various levels of education systems worldwide provides another example of datafication scholarship focused on various organisations such as technology providers and institutions such as public authorities. For example, Williamson (2017), Decuypere (2019) and Manolev et al. (2019) explore different kinds of educational technology providers, their claims, and their products. Others work together with political actors, public servants, and intermediary organisations that are involved in data-driven governance in education, (e.g. Chang, 2018; Grek et al., 2020), in science (Leonelli, 2019), and in relation to data-driven governance more broadly (e.g. Abraham et al., 2019; Micheli et al., 2020).

The review of related work presented here so far illustrates the manifold of definitions and concepts of datafication. Against this backdrop, attempts to consolidate datafication scholarship according to various analytical and socially relevant concepts have been made. Ruckenstein and Schüll (2017) propose to map datafication research in health according to three categories: datafied power, 'living with data', and data-human mediations. For the domain of education, Williamson (2018) suggests ten ways to define datafication in education, ranging from historical definitions to technical, epistemological, ontological, social, political, cultural, imaginative, dystopian, legal, and ethical concepts. For example, he suggests an imaginative and a dystopic definition of datafication: as a subject of visions possibly catalysing future real-world phenomena or as a source of anxiety (n.p.). These and other definitions of datafication showcase an important point for my argument: it is not only that different definitions of and approaches to datafication are being developed within the frames of various disciplines and fields, but these multiple definitions are being developed through diverging research practices directed at different empirical phenomena. Although this argument seems rather obvious at the first glance, it brings forth the question of how can researchers approach and reflect on the empirical and epistemological, ontological multiplicities (Mol, 2002) of datafication processes.

Bonde Thylstrup and colleagues (2019) in their edited special issue propose to focus on specific kinds of digital data addressed in research in order to better understand datafication. Agostinho et al. (2019) address datafication processes with particular attention to archives, using the analytical lens of uncertainty and propose a typology of what they call 'uncertain archives' that includes concepts of unknown and unknowable, error, and vulnerability. In contrast, Kuch et al. (2020) apply terminology of precision to describe the promises of datafication processes. Prietl and

Houben (2018) distinguish datafication scholarship according to 1) the technical artifacts and infrastructures under study (e.g. databases, protocols) as a "sociotechnical condition in operating with data" (p.12, my translation), 2) the meaning of data, and 3) cultural knowledge structure of the society. Christin (2020) proposes a framework for understanding datafication processes based on the workings, doings of data in a given empirical context, encompassing five elements such as tracking, homogenizing, triaging, nudging, and valuating. Overall, as Pellegrino (Pellegrino et al., 2019) notices,

> "[i]t is of interest that such a process, as typical in technology and media history, is subjected itself to a constant dichotomic and oppositive binary thinking […]. On the one hand, the apocalyptic register of those framing technology (in this case, data) as coming exclusively "from above", tools to exert and impose power, namely the power of surveillance and the end of privacy as we used to conceive it. On the other hand, the more "integrated" approach viewing datafication as the trigger and the field of new opportunities, benefits and progressive futures "from below"" (p.90).

Expanding on this research, with my methodological and conceptual inquiry into research on datafication, I aim to develop analytical tools that would help approaching and reflecting various multiplicities of datafication processes.

In this section, I have shown both the approaches defined in this quote as being 'from above', for example addressing the role of various technologies in datafication processes, while technologies are sometimes not considered as a part of social reality but are being called to be put in context. Others, following socio-technical approaches to defining datafication processes, view datafication processes as situated, enacted in practices of various actors, often specifically addressing the othered and the marginalised. In a critical take on critical datafication scholarship, Söderberg, in a joint work with Pellegrino and Milan (Pellegrino et al., 2019), argues, however, that detailed analytical attention to the marginalised, oppressed, and othered groups

> "engenders a 'race-to-the-bottom' that the predominant, theoretical and epistemological tenets in STS are ill suited to deal with, because those tenets cannot register cases when politically and/or epistemological weak actors are fronts for more powerful actors. New theories are needed that give guidance to inquiries into what kind of bottom-positions are really at the bottom and what bottom-positions are, on a closer inspection, much higher up in the hierarchy, when factual statements are being made." (p.97).

So, while critical datafication scholarship such as described in this and above paragraphs draws attention to the inequalities, Söderberg argues, it might also put the already marginalised groups under additional scrutiny and make them part of broader data politics rather than making their voices heard and their problems visible. Being explicit about what we—as researchers and practitioners—are talking about when talking about datafication, what kinds of actors we address, and how we are positioned in relation to these actors should also allow reflecting on the extent to which social research sometimes follows the narratives of big tech industry, for example viewing individuals as 'users' or 'producers' instead of the multiplicity of other roles, or attending to individual rather than collective agency, to technological 'fixes' rather than systemic change. In another piece exploring the role of critique for understanding datafication, Couldry (2020) also understands "social knowledge" as a "*practical understanding* of what is socially actionable" (p.1139, original emphasis). In another critical interrogation of the term 'data', Markham (2013) argues that academic knowledge production is a practice of sense-making rather than 'finding' or 'gathering' data and asks what practices are included in an inquiry that allow to overcome disciplinary histories and infrastructures and develop alternative practices of sense-making. Referring to Pink (2012), she proposes "extending the stages and senses of inquiry to include not only what is present in the most obvious 'collected' or 'collectable' sense but to also look at what is done before, during, after, and between" (n.p.). Similarly, Bechmann and Bowker (2019) in their analysis of unsupervised machine learning also elaborate on what happens before and 'behind the scenes' of a "seemingly

autonomous work" (p.1). Bonde Thylstrup and colleagues (2019) also argue for demystification of datafication required to focus on the sociocultural processes within which datafication is being enacted:

> "Our claims are more modest, attending not to the hype of data, but rather to the ways the dust of Big Data settles in mundane operations and infrastructures (Helles and Flyverbom, 2019), and how these are quietly transforming how we see, read, organise, use and dispose of knowledge" (ibid, p.3).

Alongside with discourses about knowledge production and datafication, some scholars study datafication of research processes and datafication of academia such as use of digital technologies for research or use of digital archives for academic knowledge production in social sciences more broadly (e.g. Hepp et al., 2021; Katzenbach et al., 2021). For example, practices of academic re-use of research data and open data, open knowledge, data archiving, and ongoing development of research ethics to match the ongoing datafication, are discussed and implemented (e.g. Koch, 2019). In contrast to this research, my thesis is not about datafication of academic knowledge production; *rather I am concerned with the approaches to academic knowledge production taken to understand and research other, empirical datafication processes*. While the topic of research ethics is recurringly discussed in my thesis, the former aspects such as research data management or open research data are outside of the scope of my analysis. Rather than attending to the ways in which academia itself becomes datafied and copes with such transformations, I am interested in the performative relations between conceptualisations of datafication processes such as those discussed in this section, and research methodologies applied to study these processes empirically. Particularly when digital data also become research material in studies on datafication, performative aspects of both datafication processes and research methodologies complicate our understanding of the empirical phenomena at hand. It is this performative reconfiguration of what datafication processes might mean, how they can be explored and understood that I continue to investigate in the following chapters.

As Lindgren (2020) notices, "[w]hile it has become an eternal truth, reiterated by researchers and methods teachers alike, that "the problem under investigation properly dictates the methods of investigation" […], very few of us adhere to this in practice" (p.7). To address the 'problem' of studying datafication, Lindgren proposes an ambiguous metaphor of "hacking social science": on the one hand, calling social scientists to "play around" with computational methods and on the other hand, calling social scientists to attend to datafication processes as a part of their research discipline in the first place. Other scholars (e.g. Dalton et al., 2016; Iliadis & Russo, 2016; Kennedy, 2018; Neff et al., 2017; Schrock, 2017) with their research shape a new, emerging line of thought of critical data studies or simply data studies, specifically dedicated to critical scholarship about data and datafication processes. Critical data studies build the core conceptual background of my inquiry into datafication scholarship and are the body of work to which my thesis contributes.

In this section, I reviewed a number of concepts about datafication, addressing it as an episteme, a technological process, a research agenda and a manifesto for action, and a historical continuation of other academic debates and discourses. The purpose of such listing is to illustrate how not only 'data' is an elusive term, as the quote in the beginning of this chapter states. Datafication and other currently widely used terms such as algorithms, automated systems, and AI might be even more elusive. Consider following quote by Crawford and colleagues (2014), who discuss how something

> "[…] has generated nationally funded multi-billion-dollar grant programs and tenure-track jobs across academe; it is the megafauna of the academic landscape. The rapid and widespread ascendancy of the concept attests to its significance and "stickiness" across multiple fields—it has become a "thing," despite the ways in which the term is often at odds with itself semantically and industrially" (p.1665).

In the quote from their editorial to the special section in the International Journal of Communication, Crawford et al. elaborated on the concept of 'big data'. The above argument,

however, does not lose its meaning if instead of big data we put the terms 'datafication' or 'AI' in the beginning (similar discussions on AI go back to the 1950s, see Breiter, 1995). So, if datafication can be easily exchanged with another term in the above quote, does it mean that datafication processes cannot be distinguished from mediatisation processes, quantification, or discourses about AI? With the overview provided in this chapter I showed that my answer to this question is: no. Datafication processes are defined through the 'elusiveness', relationality, and recursivity of digital data. Datafication processes render visible and tangible relations between social practices and their material, technological foundations. Datafication processes are historically, geographically, and culturally located in practices and epistemologies of the manifold of actors that enable and are affected by datafication. Datafication processes are not only practices, but also ideologies and worldviews, guiding actors' teleo-affectivities (Schatzki, 2002)—broadly understood as sets of norms and affects—to engage in practices either reinforcing datafication processes or resisting these. In relation to academic knowledge, however, datafication is only one in the manifold of the concepts that aim to describe transformational processes in our societies. In contrast to the concept of datafication, other academic discourses and fields foreground methodologies of computational techniques (e.g. studies of algorithms and automated systems, AI) or the changing role of media institutions in the society (e.g. mediatisation research), while addressing same pressing issues of justice, equality and equity, agency, power, and accountability as datafication scholarship and social sciences more broadly do.

2.4 Studying datafication processes

What are, then, methodological differences between the various ways of studying datafication processes transforming the society? In this section, I will discuss various approaches to studying digital data, datafication processes, and their implications developed in social sciences and beyond such as digital sociology, digital humanities, algorithmic research, computational social science, and critical data studies. These lines of thought have been developed in various research fields and are discussed here alongside each other as they all can be consulted in empirical research on datafication processes. As Masson (2017) argues, "data research in the humanities is necessarily interdisciplinary: it involves collaborations between scholars with backgrounds in different fields – and therefore, different views on how knowledge takes shape" (p. 26). Addressing research approaches applied to examine various datafication processes, therefore, allows to understand, how concepts of datafication are being developed and enacted, and what kinds of knowledges about datafication can be sought with the help of these methodologies. Here, methodologies and epistemologies are tightly intertwined and need to be explored in concert with one another. As Lomborg and colleagues (2020) argue in their editorial to a special issue on "Methods for datafication, datafication of methods", social sciences and, particularly media and communication research need to advance critical epistemologies and methodologies of datafication scholarship that are not only grounded in the understanding of technology, but also build on the established conceptual frameworks for understanding sociality.

> "Much research focused on datafication has concerned itself with the technologies themselves, finding innovative ways to explore how data are algorithmically processed and transforming information environments through forms of 'reverse engineering' or 'audits' as a way to highlight new forms of gate-keepers and agenda-setters (e.g. Bucher, 2012; Diakopoulos, 2015; Rogers, 2013; Sandvig et al., 2014). This has advanced understandings of datafication as decision-making systems that shape the terms of mediation, knowledge production and social exchange. At the same time, the danger of 'algorithmic fetishism' (Monahan, 2018) that drives a focus on opening up the 'black-box' of data systems as a way to make sense of digital infrastructures and social relations has led to a call for media scholars to more actively insert their long-standing engagement with the hermeneutic and

> action space between production and consumption into the study of datafication (Livingstone 2019), (re)claiming audience agency and everyday practices (Kennedy, 2018), and emphasizing the situated, contextual aspects of data as a way to understand dynamics of power (Dencik, 2019)." (Lomborg, Dencik, et al., 2020, pp. 207–208)

This growing interdisciplinary body of academic work can be addressed as critical data studies (e.g. Dalton et al., 2016; Dalton & Thatcher, 2014; Iliadis & Russo, 2016; Kennedy & Bates, 2017; Kitchin & Lauriault, 2014; Neff et al., 2017; Gerrard & Bates, 2019; Hepp et al., 2022). Research on datafication, then, can be considered as part of critical data studies. Critical data studies explore digital data and datafication processes in relation to the issues of power (e.g. Cinnamon, 2020; Lauriault & Lim, 2019; Ricaurte, 2019), politics (e.g. Baker & Karasti, 2018; Bates, 2018; Gorur & Dey, 2021; Prietl, 2019; Ruppert et al., 2017), governance (e.g. Addey & Piattoeva, 2021b; Grek et al., 2020; Introna, 2015; Leonelli, 2019; Williamson, 2016), economy, data extraction and commodification (e.g. Beer, 2018; Gerlitz & Helmond, 2013; Sadowski, 2019; West, 2019; Zuboff, 2015), civic participation and activism (e.g. Kennedy, 2018; Mattoni, 2018; Milan, 2019; Stephansen & Treré, 2019; Velkova & Kaun, 2019), bias and discrimination (e.g. Benjamin, 2019; D'Ignazio & Klein, 2020; Noble, 2018; Perez, 2019), equity, equality and coloniality (e.g. Cinnamon, 2019; Couldry & Mejias, 2019; Milan & Treré, 2019), identities and subjectivity (e.g. Kotliar, 2020; Papacharissi, 2018; Schüll, 2016), health and (self-)care (e.g. Aula, 2019; Lupton, 2012; Medina Perea, 2021), literacy and education (e.g. Bradbury & Roberts-Holmes, 2017; Livingstone & Sefton-Green, 2016; Mascheroni & Siibak, 2021), affect and emotions (e.g. Kennedy & Hill, 2018; McStay, 2020; Saifer & Dacin, 2021). While these explorations share a common critical understanding of digital data and datafication processes, they build on different theoretical concepts. For example, some authors build on the notion of 'data assemblages' addressing social, political, economic, cultural, and historical relations data create (e.g. Kitchin, 2014c) or the concept of data infrastructures, putting forward the work required to produce, process, and use digital data (e.g. Aula, 2019; Gray et al., 2018; Lai & Flensburg, 2020b; Piattoeva & Saari, 2020; Weltevrede & Jansen, 2019). Within datafication research and data studies, both conceptual and empirical research is conducted, although in their literature review Flensburg and Lomborg (2021) point out to the prevalence of conceptual research on datafication. In empirical studies, various methods can be applied ranging from dominant qualitative methods of data collection such as interviewing, to the application of computational and digital methods, for example for scraping digital data generated by social media sites' users (Flensburg & Lomborg, 2021).

Critical data studies (CDS) build a distinct line of thought about empirical datafication research encompassing an interdisciplinary body of work that critically examines socio-technical datafication processes (boyd & Crawford, 2012; Dalton et al., 2016; Hepp et al., 2022; Iliadis & Russo, 2016; Neff et al., 2017; Selwyn, 2020). Critical data studies can be characterised through the commitment to the relational, performative understanding of data as situated historically, in time, space, practices of involved and affected actors (Dalton & Thatcher, 2014). The critique, central to critical data studies is directed analytically as well as methodologically towards data and technology determinism and is enacted through research designs countering narratives of big data (ibid.), such as "data-intensive and positivistic" (Iliadis & Russo, 2016, p. 1) approaches to datafication. In their introduction to critical data studies, Iliadis and Russo (2016) outline the critical and interdisciplinary character of this literature.

> "Before and after those publications, CDS has covered a wide area of communications inquiry, including data power issues in social media, apps, the Internet, web, and platforms, but also and equally importantly statistics, policy, research, and organization. In every way that data are organized in a communicative context, CDS—as a clear call for the critical investigation of Big Data science—has coalesced around researchers ready to deploy pronounced critical frameworks in order to foreground data's power structures" (p.2).

As a result of such interdisciplinarity, three characteristics of critical data studies can be identified. First, critical data scholars seek particular kinds of knowledges about datafication processes, attending to the questions of power relations and agency, accountability, trust, identity, and subjectivity (Iliadis & Russo, 2016; see also Lomborg, Dencik, et al., 2020). Kennedy, Bates and colleagues (Gerrard & Bates, 2019; Kennedy & Bates, 2017) and Lauriault and Lim (2019) attend to the issues of data power in some cases building on work from the same-titled conference (http://datapowerconference.org/). These issues of data power present

> "discussions of inequality and injustice through three broad lenses: (1) the tactics people use to confront unequal distributions of (data) power; (2) the access to data that is most relevant and essential for particular social groups, coupled with the changing and uncertain legalities of data access; and, (3) the shaping of social relations by and through data, whether through the demands placed on app users to disclose more personal information, the use of data to construct cultures of compliance, or through the very methodologies commonly used to organise and label informatio." (Gerrard & Bates, 2019, p. 2).

So, rather than simply offering critique of datafication processes, critical data studies develop sensitivities and reflexive approaches to understanding datafication, while this reflexivity also encompasses the ways of academic knowledge production in the age of datafication within critical data studies themselves.

Second, critique in critical data studies covers both the limitations of data-driven science and is directed at the critical data scholarship itself (e.g. Bonde Thylstrup et al., 2019; Kjær et al., 2021; Lindgren, 2020), which will be discussed in more detail in the following paragraphs of this section. Finally, not only do critical data scholars apply a variety of methods and techniques including digital and computational ones, but also various methodological approaches with the focus on the research on digital data are emerging (e.g. Kubitschko & Kaun, 2016; Mützel, 2015; Velkova, 2018). Among such methods are, for example, data journeys (e.g. Aula, 2019; Bates et al., 2016; Leonelli, 2020; Medina Perea, 2021) and data walks (e.g. Jarke, 2019; van Es & de Lange, 2020) attending to the processes and politics of data movement, data diaries (e.g. Tkacz et al., 2021), feature and infrastructural analyses of social media exploring their interfaces, affordances, and interconnections (e.g. Hasinoff & Bivens, 2021; Lai & Flensburg, 2020b; Light et al., 2018).

While critical data scholars conduct question-driven research and apply a variety of methods, other researchers specifically make use of digital or computational methods and tools, for example within the domains of digital sociology, computational social science, and algorithm studies. In the late 2000s, Rogers (2009) proposed the idea of "following the medium" in research methodology and introduced a delineation between initially digital and digitalised methods. Snee et al. (2016) briefly summarise the history of digital methods, tightly connected to the development of the web from web 1.0 to the web 2.0 and social media platforms. In the early years of the web, social researchers transferred traditional methods such as interviews or survey and later ethnography into digital interviewing, online surveys, and multiple types of digital ethnography (e.g. Kozinets, 2010). According to Rogers, such methods are 'digitised'—adopted to be conducted digitally, virtually, rather than face-to-face. Other, for example computational methods are, in contrast, 'born digital' and mark a so-called computational turn in humanities and social sciences (Rogers, 2015). In their review of digital methods literatures, Carrozza and Pereira (2015) discuss that particularly within the field of media studies digital methods find wide application, for example for social network analysis (p. 223). During the Covid-19 pandemic, interest in digitised and digital methods grew significantly, as with most countries of the world entering lockdowns and reducing face-to-face social interactions, much of planned research had to be either postponed or transferred into virtual spaces. Crowd-sourced and curated reading lists on various kinds of digital methods are only one example of this increased demand and application of different kinds of digital and digitised methods (e.g. Garcia & Barclay, n.d.; Glatt, 2021; Lupton, 2021).

Digital methodologies, according to Rogers (2009) and Marres (2012), however, often focus on a snapshot of a datafication process (e.g. networks, patters of the past behaviour etc.). Further, natively digital methods are rooted in a (post-)positivist and behaviourist thinking, as these often require a big dataset for data mining and identification of specific patterns. Carrozza and Pereira (2015) enlist several research centres such as Digital Methods Initiative, Medialab-SciencesPo, Density Design, and CSISP (Centre for the Study of Invention and Social Process, Goldsmiths) within which different traditions of digital research have developed that

> "resonate with the track under 1): a) Actor-Network Theory (an approach developed within the Science and Technology Studies), b) the debate about the performativity of methods (at the intersection of Sociology and the STS); c) the opposition between the 'virtual' and the 'digital' (Media studies)" (p.225).

Similarly, digital sociology (e.g. Lupton, 2015; Marres, 2017) as a research field is also interested in more than big scale data analysis, rather attending to the relations between digital data, technology, society, and knowledge (Kennedy, 2019, p. 217 reviewing Marres 2017). An example can be found in interface methods as a way to assemble various elements of digital research including analytical techniques, data, and context of research (Marres & Gerlitz, 2016, p. 42). Additionally, interface methods also emphasise the limits of knowledge about digital data, infrastructures, and techniques of analysis. Digital humanities (e.g. Berry, 2011, 2012) are also concerned with similar questions of the relationship between society, technology, and the ways in which within this relations knowledges are produced. In contrast, the domain of computational social science is rather interested in applying a variety of methods and techniques to different kinds of big-scale datasets in order to study human behaviour (Lazer et al., 2009). Within communication and media studies, a narrower field of computational communication research is developing (Domahidi et al., 2019; van Atteveldt et al., 2019). On the one hand, it draws on the methodological innovations from computational social science, on the other hand it also addresses the challenges such a "gap between the primary purpose intended for big data and the secondary purpose found for big data [by scholars]" (van Atteveldt & Peng, 2018, p. 86). Computational social science (CSS) can be best understood through the following definition provided by Lazer and colleagues (2020), that also sheds light on the historical origins of this field.

> "We define CSS as the development and application of computational methods to complex, typically large-scale, human (sometimes simulated) behavioral data (1). Its intellectual antecedents include research on spatial data, social networks, and human coding of text and images. […] CSS encompasses language, location and movement, networks, images, and video, with the application of statistical models that capture multifarious dependencies within data. A loosely connected intellectual community of social scientists, computer scientists, statistical physicists, and others has coalesced under this umbrella phrase" (p. 1060).

Other digital researchers are also concerned with methodological questions. For example, some view digital methods as a way to overcome qualitative-quantitative divides when all research data is also digital data (e.g. Venturini & Latour, 2010). Others are also concerned with ethical questions of digital research in different contexts such as social media (e.g. Fiesler et al., 2015).

Another strain of research within algorithm studies and data science (including, but not limited to already discussed FAT studies) is concerned with data not as a by-product but a valuable output of research work, is focused on producing and disseminating data representations (e.g. visualisations), and based on the resources of already available data (Leonelli, 2019, p. 3). Another, line of algorithm studies focuses on algorithms as cultures (Dourish, 2016; Seaver, 2017), embedded socially and culturally by translating data into social categories and back (Bolin & Andersson Schwarz, 2015), what kinds of power relations do they enact (Beer, 2017), and how these power relations reiterate and reify discriminatory social relations (Benjamin, 2019; Noble, 2018). A further way to investigate datafication processes, particularly in hindsight of the issues related to the software affordances and negotiations about data representations and inscriptions in software, is

attending to software in its design, development and maintenance processes. Software studies (e.g. Chun, 2011; Galloway, 2006; Mackenzie & Vurdubakis, 2011; Manovich, 2013) and critical code studies (Fuchs, 2019) attend in particular to software as socio-technical assemblages, where data is being produced. Turning to software code and infrastructures helps to render visible, how software not only makes particular data (and relations enacted through these data) explicit, but also erases other parts (e.g. material technologies behind the software) to a significant extent. Galloway (2006, p. 320) refers to Chun (2005) as he notices how software 'hides' that what is actually material and visible: machines such as computers and further material infrastructures become intangible through the software, whereas the intangible data flows appear on the surface. In contrast to algorithm studies, software studies attend not only to the methodological problems of depicting certain data in software, but rather build on the long tradition of research about categorisations in STS (e.g. Bowker & Star, 1999) and foreground negotiations of what is represented how in software and the relevant socio-material affordances.

A critical perspective on the use of various methods for studying datafication processes highlights challenges for academic knowledge production. For example, Rieder & Röhle (2017) propose to turn away from the computation and back to different kinds of knowledges digital methods produce, identifying pitfalls of digital methods that complicate practices of knowledge production such as "the lure of objectivity", "the power of visual evidence", black-boxing, "institutional perturbations", and "the quest for universalism" (p. 112-113). Against this background, Rieder and Röhle argue that to be able to conduct digital research, a scholar requires deep theoretical knowledge not only in the empirical site of practice they investigate, but also in the methods they use (such as graph theory when using network graphs) in order to be able to draw conclusions from the application of these methods. They argue that "we need to engage in critical practice that is aware of the shocking amounts of knowledge we have stuffed into our tools" (p.123). Similarly, in his book "Data theory" Lindgren (2020) also argues that research relying on computational methods should have enough theoretical sensitivity to it (p.13). Expanding on the argument by Law (2004) about the messiness of the reality scholars aim to explore methodologically, Lindgren turns to the concepts of methodological bricolage (in his analysis also resembling the concept of triangulation (Flick, 2011) according to which methodology should be pragmatic and strategic). For Lindgren (2020) the main reason of mixing various methods is in the recursive relationship between data and society that requires both digital, computational methods and other methods well-developed for studying various societal processes and practices (p.29). In his pursue of methodological and analytical bricolage, Lindgren outlines similarities between a bottom-up, inductive approach of grounded theory and computational techniques of analysis. Arguing that both start with the same question about 'what is going on with the data', he illustrates how theoretical sensitivity in different domains of conducted research is required to be able to combine various methods and methodologies for studying datafied societies.

In contrast to this line of thought, Pasmann and Boersma (2017) question whether all the 'black boxes' of knowledge production such as computational, algorithmic techniques of data analysis, need to be opened for the researchers in order for them to understand their results. Following the arguments form so-called laboratory studies in science and technology studies, they argue that such black boxes can be made transparent in different ways:

> "if here transparency may mean to embody the functioning and dysfunction of a certain artefact: if we cannot 'know' (in an explicit sense) what algorithms do and which inaccuracies they have, can we at least embody them to such a degree that we know when to rely on their results and when to become distrustful? This would mean that they become transparent, not insofar as 'one sees what is happening', but rather in such a way that they withdraw in practice" (p.142).

As mentioned in the above sections of this chapter, however, discussing knowability of algorithmic outputs requires detail about kinds of algorithms in question: the impossibility to know is indeed

true for some of the complex ML-based algorithms, while other, rule-based algorithmic techniques can be comprehended more easily. For Pasmann and Boersma (2017), by using digital tools, researchers and users can 1) know what is unknown and therefore reflect on it respectively, and 2) know what is known without directly understanding the underpinning mechanisms. Kjær et al. (2021) make a similar argument, discussing on the example of Twitter analysis the uncertainty and partiality of social media research:

> "[a]bsences and silences in datasets thus point to the impossibility of the God's eye perspective of the traditional scholar—that is, the impossibility of the researcher as a neutral all-knowing eye. Absent data also draw attention to the ways in which digital social media research is always-already partial and shaped by the situated perspectives of the materiality of research tools and practices" (p.17).

Another challenge of digital and computational methods for studying datafication processes regards the kinds of knowledges these methods help to produce. For example, Tufekci (2014b) describes how computational politics research applies computational tools to "large datasets derived from online and off–line data sources for conducting outreach, persuasion and mobilization in the service of electing, furthering or opposing a candidate, a policy or legislation" (p. 4). As the quote suggests, applying computational methods to the studies of datafication—in the above example including online experiments—is possible due to the 'data deluge' and the availability of new kinds of digitally accessible data such as digital traces and metadata about human (political) behaviour. Acker and Donovan (2019) elaborate on this as manipulation tactics "that leverage platform features, spread disinformation by mimicking human behavior, and create swaths of networked digital traces" (p. 1598). Others, for example activist and feminist researchers point out more generative, positive examples of data analysis that benefit various communities (e.g. Cinnamon, 2020; Currie et al., 2019; D'Ignazio & Klein, 2020; Milan & Velden, 2016). So, Milan and van der Velden (2016) discuss the emergence of new epistemic cultures among civic actors that lead to an increasing use of data to the ends of the civil society. At the same time, both for practitioners and academics, issues of accessibility of data for research (e.g. APIs for scraping social media data), availability of material and financial recourses (e.g. computational power to run automated data analyses) become ever more important (e.g. Bruns & Burgess, 2016; Puschmann & Burgess, 2014).

Overall, in this section of my thesis I discussed methodological perspectives on studying datafication processes, outlining a variety of digital, digitalised, computational, and algorithmic approaches and techniques. More specifically, I outlined critical data studies as an emerging body of interdisciplinary work that brings together not only research on datafication processes in different societal domains, but also various conceptual, critical approaches to studying datafied societies. By focusing on the notion of critique in '*critical* data studies', I elaborated on the practices of academic knowledge production in times of datafication, the ways in that such a critical episteme allows to attend to the relations between data, society, and knowledge. Expanding on various growing methodological traditions such as digital methods, digital sociology, computational social sciences, I recounted major methodological benefits and pitfalls of studying datafication processes. Bringing together methodological and epistemological aspects of datafication scholarship, I argue that attending to these in tandem allows to further our understanding of the relationship between data, society, and knowledge that lie in the core of both critical data studies and other research fields such as digital sociology. The heterogenous picture of methodologies of studying datafication processes presented in this chapter serves two purposes. First, being critical to some of the mentioned lines of research, I elaborated on the ways in which these epistemological and methodological approaches to datafied societies expanded our understanding of datafication processes and fleshed out limits of knowing digitally, through digital data. Second, following the quote by Macgilchrist (2021), I aimed to showcase the heterogenous kinds of academic knowledges about datafication processes which different kinds of methods help produce.

> "By analysing heterogeneity, however, we might avoid the seductions of abstractions about the digital and find new ways of noticing and weaving together the terrible, the terrifying, the strange and the wonderful […]" (Macgilchrist, 2021, p. 665).

2.5 Construction of a field of data studies

In the previous sections I reviewed the existing concepts of datafication and methods either emerging from datafication of scholarly work and/or used to study datafied society. This overview is an empirical account of academic work in critical data studies. The place of critical data studies in academic landscape, however, is yet being determined within the manifold of other, fast developing concepts and fields discussed in this chapter so far. This section 2.4, therefore, aims to provide an overview of this body of work from the perspective of a construction of a field of (critical) data studies.

It would not be an exaggeration to suggest that each academic field goes the period of negotiation of what constitutes it and how it changes over time. Communication and media studies can serve as an example of a field that regularly enters such negotiations, not at least because of the heterogeneity of theoretical, methodological, and epistemological approaches applied within media and communication research. I draw on this example of media and communication studies for a variety of reasons. For instance, it considers itself a rather interdisciplinary or at least segmented field including a variety of theoretical, methodological, epistemological, and thematic perspectives: e.g. journalism, political communication, media activism, health, and education (see (Waisbord, 2019) for extended discussion). For example, in his analysis of communication scholarship in 1999, Craig argues that communication research cannot be described as a field due to the disciplinary fragmentation of theoretical and empirical bodies of work (p. 119). The suggested understanding of communication research as fragmented was discussed by Waisbord (2019) 30 years later as he proposes to view communication and media as a 'post-discipline'. Further, communication and media studies still can be considered a relatively young academic field and its struggles of self-definition are well-documented in multiple publications by the scholars from within (e.g. Craig, 1999, 2018). Moreover, as I have shown in this chapter of my thesis, a significant part of research that can be addressed as data studies is closely related to media studies and communication research (e.g. due to the institutional affiliations of the authors, the questions posed about digital data and datafication processes, as well as some of the seminal literature on digital data). Existing discussions on segmentation and various epistemological traditions of thought continuing within the field(s) of media studies and communication research, might also provide helpful insights for the broader discussion on the construction of a field of data studies. In this section, however, I only briefly touch upon this topic, as I pursue my methodological interest in exploring how conceptualisations of datafication in this emerging field are produced.

According to Waisbord (2019) and Craig (1999), an academic field can be understood as such when "a common awareness of certain complementarities and tensions" (Craig, 1999, p. 124) is built, meaning that a 'critical mass' of research has been accumulated that is produced, discussed, and disseminated through specific journals and associations. Lindell (2020) recounts some core aspects of an academic field by drawing on Bourdieu's (e.g. 1975) view on sociology of science. First, the author points to the recognition of political struggles within a field; second, "the common pursuit of understanding and explaining media and communication" (Lindell, 2020, p. 112). Referring to Fligstein and McAdam (2012), Lindell (2020) defines an academic field in a following way.

> "Fields tend to come into existence via 1) state facilitation, 2) the mobilisation of social actors with a common goal, 3) settlement of the field's order by key social actors, and 4) internal governance units securing the reproduction of the field" (p. 112).

Other scholars concerned with the questions of the construction of a field also point to the relevance of coherence of underlying paradigms—"a set of puzzles and a corresponding set of agreed upon methodologies for trying to solve these puzzles" (Jensen & Neumann, 2013, p. 231). As established in the previous sections of this chapter, the most important puzzle that data studies aim to solve is the relation between data, society, and knowledge that is enacted in datafication processes and the role of concepts such as power, agency, and subjectivity in these relations. The underlying and widely accepted principles of data relationality, recursivity, rejection of neutrality of technology and data, as well as processual understanding of datafication have been established in the early work on data studies developed at the time when Mayer-Schönberger and Cukier (2013) broadly introduced the concept of datafication. These seminal conceptual works, for example by boyd and Crawford (2012), Kitchin (2014c), Iliadis & Russo (2016), Kennedy and Bates (2017), and others advanced conceptual and critical understanding of digital data and datafication processes.

More recently, empirical research on datafication has been published increasingly. As Flensburg and Lomborg (2021) discuss in their recent systematic review of datafication research, among the central topics within data studies conceptual work on datafication is still dominant (p.11). In regard to methodologies applied in data studies and other related fields, in the previous section I outlined the heterogeneity of epistemologies, approaches, and techniques used to explore and explain datafied societies. As Flensburg and Lomborg (2021) point out, datafication research, however, is predominantly conducted by qualitative scholars from the broad domain of social sciences (p. 8 & p. 12). Only a small part of contributions the authors reviewed has been found to apply digital or 'data-driven' methods (p. 12). Additionally, according to Flensburg & Lomborg (2021) much literature about datafication addresses users and the social implications of datafication or technological aspects of datafication processes, while the latter strain of empirical research is rather scarce in relation to empirical studies of the social implications of datafication (ibid.). Surprisingly, according to Flensburg and Lomborg (2021),

> "[d]atafication literature seldom reflects critically on the use of the concept and how it relates to the chosen objects of analysis, methods, and empirical source materials. If we are to stick with datafication as a useful conceptual lens onto contemporary data developments and uses in society, our analysis raises questions for the ability of research to cumulate systematically to the mutual enrichment of the fields that have stakes in the topic of datafication." (p.14).

These findings suggest that while there is methodological consistency in the emerging field of data studies, more empirical work and methodological heterogeneity are required in order to advance our understanding of datafication processes and what 'datafication' as a concept might mean empirically for different actors. The above quote also illustrates how datafication scholarship, while attending to media, technologies, and data, does not follow a media- or techno-centric approach. Rather, a socio-technical view on datafication processes aims to close or at least narrow the gap in the binary between the technological and the social and develop research agendas directed towards a more complex picture where humans and other actors co-exist and co-construct with one another what we call 'society' that is in situ datafied.

Further following the list of elements that help the construction of a research field, academic associations, conferences, and journals as spaces for research collaborations and dissemination need to be discussed. There are multiple academic associations and their core conferences for social science research that harbour datafication scholarship such as Association of Internet Researchers (AoIR, https://aoir.org/), or Fairness, Accountability, and Transparency in Machine Learning (FATML, https://www.fatml.org/); their focus is, however, either broader than that of data studies or narrower. Besides FATML a number of networks around critical algorithm and AI research have been established in recent years such as the AI now institute (https://ainowinstitute.org/), DAIR (https://www.dair-institute.org/) and other. Among the conferences specifically dedicated to some of the core questions of data studies are the Data Power conference (http://datapowerconference.org/), the Data Justice conference organised by Data

Justice Lab at the University of Cardiff (https://datajusticelab.org/data-justice-conference/) and the "Agency in a datafied society: Communication between and across humans, platforms and machines" (https://www.uni-bremen.de/zemki/veranstaltungen/tagungen/agency-in-a-datafied-society-communication-between-and-across-humans-platforms-and-machines) conference held 2021 at the University of Bremen. Further work also emerges at the intersection of social sciences, humanities, and computer science, human-computer interaction, notably published at such conferences as ACM Conference on Human Factors in Computing Systems (ACM CHI, https://sigchi.org/conferences/conference-history/chi/) and ACM Conference on Computer Supported Cooperative Work (CSCW, https://dl.acm.org/conference/cscw). In addition to the spaces for research collaboration and dissemination, to develop a 'critical mass of research', enough scholars must have expertise in the emerging field, also meaning that appropriate university programs, research positions and institutions are established (Waisbord, 2019). At the time of writing, some academic positions, including professorships have been advertised that included 'critical data studies', alongside with, for example, digital methods and studies of digital society, in the scope of topics such new professorships should cover. This indicates the growing importance of data studies as both a research topic and a field. In regard to university programmes, (critical) data studies are not yet as established as, for example, digital sociology (although some refer to digital sociology as a 'field in formation', s. Gregory et al., 2016, p. xx) or computational social science as students willing to study the latter can do so both at undergraduate and post-graduate (e.g. master's) levels in many universities, at least in Europe and North America.

Finally, negotiations about what constitutes quality and rigor in the emerging field of research take place in which the possibilities, limits, and perils of scholarly work in an emerging research field are elaborated. In that sense, ontological and epistemological debates about disciplinary boundaries and expertise connect to the methodological discussions of objectivity and rigor—ontological politics (Mol, 2002). Sismondo (2010) further discusses how scientific credibility is supported by standardisation of research procedures which, in turn, promise higher stakes in application for funding (p.138). He portrays the historical development of academic credibility from the 'absolute' objectivity to expertise and the ways that expertise can be gained (p. 139). Knorr-Cetina (2002) discusses that research practices need to be understood from a constructivist perspective, they are informal, political, embodied, rely to some extent on tacit–-implicit–-besides the explicit knowledge. Building upon such social constructivist arguments about scientific knowledge, Haraway (1988) connects the discussion of academic objectivity (as opposed to rigor) with the discourses of power and the question of who holds that power in academic fields (p. 577). However, as Haraway (1988) states, obscuring objectivity to power does not help to produce knowledge about the social world: "[b]ut we could see some enforceable, reliable accounts of things not reducible to power moves and agonistic, high-status games of rhetoric or scientistic, positivist arrogance" (p. 580). Instead, it rather perpetuates the dominant view on the world. According to Haraway, a way to produce objective (and also rigorous) accounts not reduced to academic and disciplinary power relations is to engage in production of embodied, situated, and partial knowledges. Situated knowledges should offer a "partial perspective" (p.583) that helps to reconstruct "how to attach the objective to our theoretical and political scanners in order to name where we are and are not" (Haraway, 1988, p. 582). In contrast to tacit knowledge mentioned in the beginning of this section, situated knowledge does not refer to the ways in which people, including researchers, apply their knowledge or the extent to which they can articulate it; rather, situated knowledges put forward practices of reflection about what is known and why, what is othered in this knowledge, and how this knowledge is related to the social reality of the knower. Recounting which social, technological, material, and emotional issues comprise what we understand as academic knowledge and taking stand to it, *positioning* oneself to it, means to produce situated knowledges. Haraway points out that in order to produce situated knowledges, we need to turn our attention and explicate the ways in that our bodies, methods, devices, as well as theoretical concepts

that we use for academic inquiry, produce a partial view about the studied phenomena. As I briefly discussed in this section of my thesis attending to some works from the sociology of scientific knowledge, these various aspects are inherent elements of practices of academic knowledge production. In my methodological and conceptual inquiry into research on datafication, I explore empirically, how studying these aspects of datafication scholarship allows a better understanding of mutually co-producing relations between research methodologies and empirical conceptualisations of datafication processes.

2.6 Situating knowledges about data and datafication processes

With the expansion of datafication to all societal domains, several challenges for production of situated knowledges emerge. First, research needs to deal with an apparent opacity of 'mediating devices' people, including academics, use to produce knowledge about datafied societies like surveillance systems or artificial intelligence-linked systems (Haraway, 1988, p. 581). When attending to datafication processes empirically, however, reflecting on the role of digital devices in the academic knowledge production and situating concepts about datafication in empirical realities may become a more complex task. In the previous section I introduced the main challenges of academic knowledge production in times of datafication. Issues such as theoretical underpinnings of the automated, computational research techniques, repurposing of digital data gathered by commercial actors for research, biases in the data-driven methodologies, and ethical challenges to the use of unconsciously left digital traces are among the central ones. Furthermore, multiple scholars have attended to the question of what can be known or unknown in research based on computational, digital, and data-driven methods, particularly when the scholars themselves have little resources changing algorithmic techniques, for example when using specific research tools or relying on APIs provided by commercial companies for data scraping. Second, Haraway (1988, p. 584) warns against knowledge production solely from 'subjugated' positions, arguing instead for embracing the multiplicities of the positionings that scholars can produce (p. 586). Third, Haraway (1988, p. 593) interrogates the notion of agency in relation to the practices of knowledge production. For Haraway, an object of study about which situated, partial knowledge is produced, is not necessarily a passive, inactive *object*. Instead, she argues for recognising agency.

> "The point is paradigmatically clear in critical approaches to the social and human sciences, where the agency of people studied itself transforms the entire project of producing social theory. Indeed, coming to terms with the agency of the "objects" studied is the only way to avoid gross error and false knowledge of many kinds in these sciences. But the same point must apply to the other knowledge projects […]" (p. 592).

In datafication scholarship, agency is among the central analytical concepts under investigation. The questions about who can exercise agency, the extent to which agency can be attributed to non-human actors such as technologies broadly and data or algorithms more specifically, the extent to which users of such technologies have leeway for their agency or are at the mercy of technological agents and technology providers, and how to hold tech industry to account for their activities are among the most pressing issues that datafication scholars explore in their research. The latest calls for more empirical data studies of the everyday practices and for recognising 'users' as knowing and agentic, resonate with this final aspect of situated knowledge production.

Such critical questions and attention of critical data studies to the relations of power enacted through and within datafication processes puts forth the notion of 'critique'. While critique as a theoretical and epistemological foundation is in the core of critical data studies, some scholars propose to address this body of work as data studies (e.g. Kennedy & Bates, 2017), without mentioning 'critical'. In his seminal text about scholarly critique, Latour (2004) "bring[s] the sword of criticism to criticism itself" (p.227) by outlining two common critical scholarly positions, briefly described as debunking beliefs as projections and making attributions to practices (p. 237-238).

Latour's answer to his title question on why critique has run out of steam is that both these strains of critical thought are not only always giving correct interpretation in relation to the positions of the critics themselves, but also disempower both the beliefs and the attributions. For Latour, such critique is a *matter of fact* analysis that gets further and not closer to the realities it aims to discuss. Rather, critique should be an *inquiry directed to the matters of concern*—arenas for gathering the multitude of elements required for the reality to be enacted.

> "The critic is not the one who lifts the rugs from under the feet of the naïve believers, but the one who offers the participants arenas in which to gather. The critic is not the one who alternates haphazardly between antifetishism and positivism like the drunk iconoclast drawn by Goya, but the one for whom, if something is constructed, then it means it is fragile and thus in great need of care and caution" (Latour, 2004, p. 246).

Thus, Latour argues for more positive, generative critique that produces "more ideas than we have received" (p. 248). While research talking about its objects of inquiry in a matter-of-factual way detaches this object of inquiry from its and researchers' social realities, matters of *concern* shed light on the intertwined relations between research, its objects, and other societal domains. Matters of concern, thus, are bound to highlight researchers' own positionings in their research field and in the society. Scholars working in feminist traditions of thought forwarded this argument further, arguing that matters of concern, as well, are still detached from actual empirical phenomena, and introduced the notion of "matters of care" (Puig de la Bellacasa, 2011, 2017). It expands on the feminist concept and ethics of care, that has been applied across various research disciplines and fields—from gender studies to political science and health—in order to underscore affective, material, normative, and ethical dimensions of social relations (e.g. Lindén & Lydahl, 2021; Mol, 2008; Tronto, 1993). The notion of matters of care proposed by Puig de la Bellacasa (2017) draws attention to the emotional, embodied, intersectional, and ethical aspects, also required to enact realities and situate these beyond facts and concerns. That resonates with the critique of critical data and algorithm studies offered through the concepts of data justice or data colonialism discussed in the previous sections, such as an overemphasised attention to the technical, the methodical, and the Western. Finally, some of the scholars advancing the emerging field of critical data studies such as Kennedy and Bates (2017) also leave out the notion of 'critique' in the description of the research process to its content. In line with and *caring* about these arguments, in the following chapters I continue my inquiry into the development of *data studies* as an academic field.

In sum, within data studies their role as a field of research and their position in relation to other fields are currently being negotiated in conceptual and empirical literature. Alongside with the pressing empirical questions on the relations between data, technology, subjects, and society and the implications for our understandings of power, agency, and subjectivity, data studies negotiate how academic and practical knowledge can be produced and in what ways data studies shape, complicate, and trouble such knowledge production. While there is a breadth of conceptual approaches to datafication, recent literature reviews (e.g. Flensburg & Lomborg, 2021) show that more empirical research on datafication is needed that would continue the negotiations about more situated, partial perspectives on datafication. With my thesis I aim to further the discussion on data studies as an emerging research field and, following the research agenda for empirical datafication scholarship set by Flensburg and Lomborg (2021) explore empirical datafication scholarship in more detail. I also expand on the argument that datafication scholarship requires more methodological reflection and, in my empirical study, examine methodologies applied to study datafication processes in order to develop a vocabulary and a heuristic for data studies that is grounded in its inherently critical, relational view on data and datafication.

3 Research methods as objects of inquiry

Susan Leigh Star has famously asked in her book about knowledge production in science and technology studies, "[c]ui bono? Who is doing the dishes? Where is the garbage going? What is the material basis for practice? Who owns the means of knowledge production?" (Star, 1995, p. 3). In this chapter, I address the role of methodologies in knowledge production, and how research methodologies themselves have been addressed conceptually. By turning to these questions, I aim to explore, what are and 'who owns the means of knowledge production' in empirical datafication research. An increasing number of scholars propose new methodologies including creative, inventive, and artistic methods among others. These should add to the traditional research designs, struggling to grasp the processual character and fluidity of datafication processes (Kitchin, 2014c; Law, 2004, p. 4). Usually, empirical studies outnumber theoretical articles in some of the most renown journals in communication and media sciences (Waisbord, 2019) and present a plenitude of methods and methodologies practiced and enacted in various empirical fields. Specifically to datafication scholarship, conceptual contributions prevail while empirical work is increasingly being published on the topic (Flensburg & Lomborg, 2021). At the same time, special issues' and handbooks' authors and editors reframe a method from a mere 'tool' to conduct research to the object of inquiry, for example addressing the role of methods in research in general, and in research on datafication in particular, or by discussing new, 'innovative' methods and how they change research practices (e.g. Hand & Hillyard, 2014; Kara, 2020; Koro-Ljungberg & Mazzei, 2012; Kubitschko & Kaun, 2016; Lomborg, Dencik, et al., 2020; Lury, 2020; Lury & Wakeford, 2012; Marres, 2017; Snee et al., 2016).

The growing number of authors engaging with research methods in the light of datafication only manifest the rising relevance of methodological issues within various research communities. The development of digital methods and datafied research facilitated an academic debate around the questions of what these new methods do and how their doings—methodological practices—are changing research processes. So, method is not only a technique for conducting empirical studies, but also an object of inquiry in itself, as it requires critical interrogation and conceptualisation. To understand current academic discourse on research methods as an object of scientific inquiry, I first elaborate on the evolution of research methods. A view from the historical perspective sheds light on methodological and related epistemological, ontological concepts, underlying each method and helps to contextualise current discussion on the methods according to their philosophical background. Specifically, I pinpoint the changes driven by the growing interdisciplinarity of research, the developments of the new materialist and relational ontologies as milestones in the ongoing discussion on the role of methods in studying datafication.

One example for the development of these ideas is a methodological debate called "the social life of method" (Law et al., 2011; Ruppert et al., 2013; Savage, 2013). In this chapter, I attempt to connect various methodological approaches to studying datafication and data to the challenges identified through the social life of methods debate. Finally, I attend to the critique of the social life of methods and introduce parallel methodological developments, focused primarily on different ways of engagement with research practices at different empirical sites. With my analytical focus on methods as objects of inquiry, I use these critical discussions to identify categories suitable for an operationalisation of research method as an analytical concept.

Defining method through multiple categories suggests that we cannot treat it as a single "thing" or practice. Therefore, I apply a theoretical approach of methods assemblage (Law, 2004; Mol, 2002) to interrogate how the entanglements of research method are brought together and (re-)negotiated in research practices and products, i.e. published articles. A methods assemblage is constituted both as a part and as a consequence of the research process through the interactions between human researchers, their methodological tools and artifacts, and human or non-human research objects. A methods assemblage helps a researcher to find what they are looking for within their specific objects of study, operating within specific philosophical frameworks. The methods assemblage, therefore, underscores the performative character of methods. The concept of methods' performativity (Barad, 2007; Law, 2004; Mol, 2002) denies the dualism of the researcher and the researched. This understanding of performativity is central to the notion of a methods assemblage (Law, 2004). While academic rigor and 'objectivity' of scientific knowledge are in the foreground of the performativity concept, it further emphasises the situatedness of that knowledge. As various research and methodological practices are in the core of what we conceive as methods and, respectively, methods assemblages, I finally discuss the work and decision-making of datafication scholars in their empirical research. The aim of this chapter is to review related literature in which research methods are treated as objects of inquiry themselves and introduce the notion of methods assemblages as a concept for addressing practical, situated, and reflective work required for datafication research.

3.1 On methodologies, epistemologies, and ontologies in academic knowledge production

There are long-standing traditions of thought concerned with academic knowledge production, addressing how academic knowledge is developing over long periods of time, how changes in the knowledge production come to be (Kuhn, 2020 [1976]), and what particular research and epistemic practices allow production of such knowledge (Knorr-Cetina, 2002; Latour & Woolgar, 1986). This body of work is also addressed as sociology of scientific knowledge. In its core is the idea of social construction or shaping of knowledge production, and mechanisms of uncertainty and closure relevant for understanding how development of that knowledge is carried out (Pinch, 2015). Kuhn (2020 [1976]) argued that scientific knowledge is not accumulated, but rather changes following scientific revolutions. These revolutions bring about a shift in paradigms—the central conceptions agreed upon in an academic discipline, including theoretical and methodological rules. Others, like Fleck, introduced similar ideas earlier than Kuhn, e.g. discussing 'thought collectives' (Fleck, 2019 [1935]), or provided critique of Kuhn's work (e.g. see Bourdieu, 1975, p. 22 for a comment on 'idealist philosophy'). While Kuhn's definition of a paradigm remains debatable and, as Kuhn himself noted, draws on the history of natural, experimental sciences (Kuhn, 2020 [1976], p. 171), it was one of the starting points for further discussions about the ways in which scientific knowledge is produced and what does is mean for science. Some of the scholars, like Pickering (2012), however, discuss how these arguments can also be applied to other disciplines than natural sciences such as early AI research. In data studies, Kitchin (2014b) drawing on Kuhn's work suggests that there is a "possibility of a new research paradigm across multiple disciplines" (p. 3) being under way, as digital data become pervasive in practices of knowledge production. Despite Kuhn's ideas being widely debated, as Sismondo (2012) puts it in the special section of the Social Studies of Science Journal celebrating 50 years of Kuhn's seminal book, "Kuhn challenged dominant popular and philosophical pictures of the history of science, rejecting formalist accounts in favour of attention to the cultures and activities of scientific research" (p.415). Broadly drawing on these arguments, other scholars such as Latour and Woolgar (1986), Knorr-Cetina (2002, 2016 [1984]), Pinch and Bijker (1984), Collins and Pinch (2012 [1998]), MacKenzie and Wajcman (1999) have discussed the role of science in the society, its relations to technology and technological

developments, and how sociality of technology is constructed, not at least, through science and vice versa. For example, according to Knorr-Cetina (2002), attending to academic cultures—which she addresses as epistemic—allows exploring the informal and the 'cultural' aspects of academia: power relations, practices of interpretive research work, embodied experiences and tacit knowledge of individual scholars, as well as collective practices and values through which these epistemic cultures are being negotiated. Describing academic collectives, Knorr-Cetina (2002) argues, we should focus on the epistemic practices and the ways in which research is being conducted, besides its disciplinary context (p. 12). Academic knowledge, then, becomes a *product* generated and available in certain settings (ibid., p. 17). Viewing knowledge production in this way raises questions about patterns such knowledge production follows, highlighting multiple social, political, and other aspects in the construction of scientific arguments and "facts" (Latour & Woolgar, 1986). This focus on science entangled in other societal processes, however, has also been criticised. For example Fuller (2012) argues that Kuhn and Latour neglect "the normative sensibility that lay behind the desire to keep science, in some sense, 'autonomous' from the rest of society" (p. 431). Turner (2012) contends that

> "[t]he hard work […] is to epistemologize the social, to understand the elements of belief and belief formation, which inevitably depend on our knowledge of others and of institutional routines […]. And this will force us to ask the politically uncomfortable questions we find so difficult, questions about when to believe experts, about whether and when the consensus formation processes of science can be relied on, and to face our prejudices with hard questions" (p. 479).

The questions raised here, however, are not so different from the arguments criticised scholars of scientific knowledge production make and imply not at least what can be addressed as research politics: the roles of academic institutions, hierarchies and relations of power, normativities and expectations from certain scientific endeavours as well as their costs that altogether partake in the formation of consensus in science (also see Bourdieu, 1975 for discussions of the 'stakes' in the struggles in an academic field). In STS, practices of reflection on research processes have been addressed as helpful for making sense of these complex relations (see Ashmore, 2015). As Cohen and Galusky (2010) put it,

> "[t]o some degree this is a turn toward the self and personal reflection as a way to communicate research and illustrate embodied conflicts. Importantly, however, it is not a turn away from the world or a slipping into some inflated sense of self-importance" (p. 2-3).

In my thesis, this practice of reflection on research process becomes one of the central sources of analysis: it is the scholars conducting empirical research on datafication, whose reflection on their own study design, methodological choices, and practices, is driving my analysis presented in the chapter 6 of my thesis.

Similar to Kuhn's book, many studies concerned with academic knowledge production refer to research in natural sciences and scientific laboratories (from physics to biology and medicine). One of the specific characteristics of such laboratories can be well captured by what Knorr-Cetina (2002, p. 48-49) calls a "dissecting room". In a laboratory, research takes place in such a dissecting room, at least somewhat separated and detached from the (social) reality of the empirical phenomena under study. As the overview of empirical research on datafication, provided in chapter 2 of my thesis, has shown, in a somewhat similar manner—'somewhat' here being crucial—empirical research on datafication also seeks to find its dissecting rooms in graphical user interfaces, algorithmic code, PR and marketing materials of technology providers, or technology design and development documentation. It is, then *somewhat* similar to the dissecting rooms of experimental sciences even though the settings of these two kinds of dissecting rooms are different: in empirical datafication research they may not (always) be created through actual, physical walls and research machinery such as microscopes and Petri dishes, but rather these rooms are being created through the application of certain research methods and tools, taking certain elements of

datafication processes out of their social reality and putting them on the researchers' desk (or their monitor, for that matter). My use of theories of practice (as described in chapter 4), thus, also draws back on these early studies of academic knowledge production which focused on specific research and epistemic practices/activities in scientific communities and laboratories. This practice-theoretical perspective draws my empirical attention to methods assemblages. As I will show in this chapter, methods assemblages can be considered as practices of weaving together various elements of academic research (the researcher, the researched, the specific research procedures, all situated institutionally and materially). In this sense, by attending to methods assemblages analytically, I explore how researchers conducting empirical studies on datafication detach, but also how they enact new social realities in which these studied datafication processes are then being re-situated. With that, I also follow a call by Diaz-Bone and colleagues (2020) for the sociology of social research that is particularly attentive to the role of digital data in it. Before continuing this argument, however, I give a brief historical overview over research methods in social sciences. This overview is required to turn our attention back from the discussions of scientific knowledge production in natural sciences starting from 1970-80s to the developments in social sciences happening at the same time. It shows why my argument is not about 'the method', but rather about a complex assemblage of various elements held together through the research practices. Such an assemblage also includes the researcher, an object of inquiry and the field where research is conducted and is interwoven with respective epistemologies—approaches to define knowing, and ontologies—approaches to define being, applied by the researcher in a particular study, and methodologies—a strategy and a setting guiding researchers' choices. In my thesis, I give preference to the term 'methods assemblage', as it underlines the tying together of many different, otherwise possibly disconnected elements for the purpose of academic research. Considering similarities between the terms 'methods assemblage' and 'methodology', when I need to use an adjective, I write 'methodological'.

I attend to the historical overview from the perspective of underlying philosophical assumptions in their relation to the methodologies and methods (see for detailed discussion e.g. Denzin & Lincoln, 2018). Various scholarly communities conducting research on digital data can be addressed within the frameworks ranging i.e. from (post-)positivist, according to which methods can capture the reality as it is; interpretive, according to which methods provide a way to understand the reality, to new materialist, according to which methods are actively reconfiguring research processes. While these and other (e.g. critical) understandings of methods currently exist, at least to some extent, in different research communities, they also reflect the historical evolution of methodological thought. The 'scientific' method, characteristic to positivism, strongly prevailed until 1980s. This understanding of method is backed through the idea that the goal of science is to unveil the facts about the world. For that, some analytical tools—methods—are required. Researchers are 'objective', 'impartial' observers of the natural world. To make inquiries about that world means to search for a correct answer, while methods are not meddling with the world but allow to capture it 'as it is'. Accordingly, 'hard' methods based on quantitative measurement and representativeness of the research results were flourishing. This worldview and the leading position of positivism have been contested in 1970-80s in advantage to the post-positivism, acknowledging some constructivist ideas about the role of people in general and researchers in particular in the making—construction—of what we refer to as the reality. That development leads to further flourishing of qualitative research, however the distinction between 'hard' quantitative and 'soft' qualitative methods, also known as 'methods wars' (see for detailed discussion e.g. Denzin & Lincoln, 2018), is preserved (Onwuegbuzie & Leech, 2005). The quantitative research data is understood as objective representation of the reality, while the qualitative research and data it produces is seen as a 'subjective' and not representative of the world. These differences do not only refer to the diverging theoretical underpinnings, but also to the methods used to conduct quantitative and qualitative studies. Within the qualitative methods community, the position of an interpreting researcher consolidated.

An interpretation of an object of inquiry rather than discovery of single 'truth' about that object is in the core of interpretive research. In this period, methods' development is mostly allocated within either qualitative (constructivist, interpretive) and quantitative (rather positivist) approaches and underlying paradigms. The 'pragmatic' approaches mixing these two slowly developed since the 1990s (Denzin & Lincoln, 2018, p. 5; Onwuegbuzie & Leech, 2005), and continue to play a central role in the current methodological discussions (e.g. Watson, 2020). The subsequent conflict between proponents of mixed research and the so-called methods 'purists' put the questions of ontology and methodology in the foreground, therefore further making methods to the object of inquiry. It means, not only particular techniques and procedures for data collection and analysis, but also the questions about what kinds of answers do different kinds of method help acquiring were in the foreground. These issues, however, also have a long history, not at least in the philosophy and sociology of science. For example, Feyerabend (2010 [1988]) in his seminal piece "Against Method" questioned whether and how 'a' methodology can provide any principals for 'a' scientific inquiry: he argued instead that in science, there is a multiplicity of inquiries each of which requires its own, specific, tailored methodological approach. Ian Hacking in the preface to the Fourth Edition of "Against Method" (2010 [1988]) reflects on the infamous argument "anything goes" formulated by Feyerabend: Hacking writes

> "we must emphasize that Feyerabend never meant for one minute that anything *except* the scientific method (whatever that is) 'goes'. He meant that lots of ways of getting on, *including* the innumareble methods of the diverse sciences 'go'" (p. xiii, original emphasis).

Faced with the social reality, researchers develop exploration techniques focused on particular empirical aspects, and following particular goals (see for the critique of the "knowing capitalism" Burrows & Savage, 2014; for some examples of the political influence of methodological innovations Savage & Burrows, 2007).

Some scholars argued, however, as the empirical social world changes, the methods of study also require change (Law, 2004; Marres, 2017; Rogers, 2009, 2019; Savage & Burrows, 2007), otherwise these methods run the risk of oversimplification or simply re-producing normativity (Koro-Ljungberg & Mazzei, 2012). Especially with the ongoing digitisation and datafication, the discussions about the kinds of knowledges produced by different methods became more pressing. As described in the previous chapter, researching data and datafication challenges the representational and referential approaches to academic inquiry. Thus, most recent methodological developments are situated in nonrepresentational (Vannini, 2015), new materialist (Fox & Alldred, 2018; Moor & Uprichard, 2014; Nordstrom, 2015), postqualitative (St. Pierre, 2018, 2019), and posthuman (Barad, 2007; Braidotti, 2019) approaches. Alongside with methodologies embedded in these varying concepts, there have been calls for antimethodologies (Nordstrom, 2018) or 'against method' (Manning, 2015). All these methodological developments are mainly situated in the former communities of rather qualitative scholarship. Parallel to the methodological history in the Western strain of thought, Indigenous traditions developed rather separately (see e.g. Rosiek et al., 2020 for critique and reasons to why Indigenous research still often lacks western scholarly attention). As Indigenous scholars notice, Indigenous ontologies have various similarities to the new materialist approaches, foregrounding non-human agency and performativity/reciprocity in the relations between both human and non-human actors (Rosiek et al., 2020). In contrast and in addition to the western new materialist approaches, Indigenous scholars provide pathways to overcome the multiple dualities of research and set out to produce different kinds of knowledges. For example, Indigenous studies illustrate a development beyond the point where the non-human agency per se is the core topic of inquiry. Rather, attention is turned to the particularities (the "how?") of knowledge production in the complex assemblages, entanglements of humans and non-humans underscores the necessity to explore these questions as well (Rosiek et al., 2020).

Having described the development of qualitative methods in social sciences, a peek into the emergence of current quantitative line of research is noteworthy. So, quantitative methodologies

evolved to the evidence-based approaches, currently often represented through computational methods for big data analysis based on identification of patterns. In contrast to the new empiricisms, partially denying the role of methods in the research process, the evidence-based quantitative research put methods and not theory or experience at its core. So, in his polarizing article, the former Wired editor Anderson (2008) proposed the dismissal of theory in favour of data-driven methods of knowledge production. Therefore, according to Denzin and Lincoln (2018, p. 5) the current methodological discourse can be described as follows. On the one hand, there is an opposition between evidence-based and integrated interpretive and new materialist lines of thought, as it is visible in the critique of 'native' digital methods, situated within critical data studies (Dalton et al., 2016; Dalton & Thatcher, 2014; Iliadis & Russo, 2016; Kitchin & Lauriault, 2014). On the other hand, there is an opposition within the qualitative studies, where the roles of scholars, methods and objects of inquiry are shifting between representational, constructivist and nonrepresentational, new materialist approaches. The latter situate the researchers, their tools, and objects of study within the empirical field and underscore the situatedness of the research (Piattoeva & Saari, 2020). Despite the ontological and theoretical differences within these reflexive, critical approaches, they share the common understanding of the reality and qualitative research in particular as messy, changing, and material processes. It is that strain of thought that I follow in my research, understanding methods as assemblages of people, non-human entities, and practices that tie these together for the purposes of academic inquiry. While the role of methods assemblages in contexts outside of academia is briefly reflected in the analytical chapters of my thesis, my primary focus is on methods assemblages for academic knowledge production.

Ontologically and epistemologically, I situate my research in relational ontologies (Barad, 2007) that position researcher and their practices of research and knowledge production within the empirical field and not outside of it. My goal in providing this short and, by far not exhaustive overview is, first, to provide a historical perspective on how and why research methods and methodologies have been explored as objects of inquiry. Second, I aimed to show the development of methods from a mere tool of 'capturing the reality' to a performative device intertwined with the research site in a complex assemblage. In the following, I attend in more detail to the so-called "social life of methods" (Law et al., 2011) debate to exemplify the ways in which methods co-produce sociality.

3.2 Research methods in academic practice and knowledge production

One of the best ways to learn about research methods in academic practice and knowledge production is through methods handbooks and textbooks. The methods handbooks and textbooks usually address as their readers not only peer scholars but also students of different levels. In their examination of research methods, such handbooks, therefore, are primarily focused on the question of how to conduct certain studies in certain empirical situations (Boellstorff et al., 2012, p. 2; Hargittai, 2020, p. 1). Such handbooks and textbooks shape the role and place of research methods in academic practice and knowledge production as they build the background of methodological expertise for current and future social science scholars. While there is an abundance of books elaborating on various kinds of digital methods (e.g. Hargittai, 2020; Rogers, 2019; Salganik, 2017; Snee et al., 2016; van Atteveldt & Peng, 2018), or on techniques of data science, books specifically addressing methodological reflection and challenges of data studies, grounded in critical, relational approaches to datafied societies are only emerging (e.g. Kubitschko & Kaun, 2016; Lury, 2020).

The common theme of multiple handbooks on research methods in the digital age and/or datafied society are the opportunities new, digital, computational research techniques open for social sciences, often in combination with other, well-established techniques (e.g. Hargittai, 2020; Salganik, 2017; Salmons, 2015; Snee et al., 2016). Although there are handbooks that provide examples of both qualitative and quantitative methods (e.g. Hargittai, 2020; Salganik, 2017), many

authors focus on one type of method like ethnography (e.g. Boellstorff et al., 2012), qualitative methods broadly defined (e.g. Hand & Hillyard, 2014; Salmons, 2015), digital methods (e.g. Rogers, 2019; Snee et al., 2016), or computational methods (e.g. van Atteveldt & Peng, 2018). Others provide methodological guidance for scholars from particular empirical fields such as, for example, digital migration studies (Sandberg et al., 2022). In relation to research methods for studying datafied societies, several issues are discussed across the different kinds of methods textbooks: research ethics, implications of the application of certain methods such as unwanted visibility or surveillance (e.g. discussed in the introduction to Sandberg et al., 2022), acquisition of research participants (e.g. discussed in the introduction to Hargittai, 2020), or issues of access to data (e.g. discussed in the conclusion to Snee et al., 2016).

The usual structure of a method textbook for studying digital/datafied societies, for example through online cultures (e.g. Salmons, 2015) or, broadly, Internet includes an overview over relevant methodological traditions, discussions about research ethics, and a list of techniques of data collection and analysis, finally complemented by advice on writing up and reporting results. In addition to practical advice about specific methods and techniques, methods handbooks also provide guidance, particularly for students and junior scholars, in developing research questions, identifying objects of study, and developing research designs for their projects. The issues of sampling, construction of questions and analytical instruments or categories become central while the empirical site discussed in the examples of methods' application primarily serve as an illustration. Methods handbooks, alongside with academic journals focusing on research practices such as Qualitative Inquiry or Qualitative Research, become the spaces in which methodological reflection takes place, often separated from other theoretical and conceptual discussions. This allows the authors to address methods as objects of inquiry through a number of reflective questions such as: what are particular methods, what quality criteria can be applied to them, how do I know that this method is appropriate tool for engaging with my research question. At the same time, such methodological reflections build a specific genre of text, foregrounding the role and performativity of research methodology (see e.g. introduction to Hargittai, 2020). Outside of the scope of methods handbooks, however, methodological sections in academic articles about datafication processes sometimes come too short, arguably in favour of more detailed reports on conceptual grounding or empirical results (Kennedy et al., 2020, p. 44). This is also backed by the attention to the issues of validity and reproducibility of methods also tied to "the audit contexts of work in contemporary academic cultures" (Garforth, 2012, p. 266).

> "What counts in the assessment of science is research understood exclusively as active doing. Modes of knowing that foreground embodied expertise and interaction with no object/ifiable outcome—outputs in the language of research assessment; knowledge claims or "facts" in the language of Latourian STS—are difficult to defend or even to articulate. They are literally invisible and discursively absent. The black-boxing of cognition in recent STS does not necessarily contribute to the audit logics of contemporary research cultures" (ibid., p. 280).

While these issues are crucial for rigorous academic work, reflections on the notions of methods performativity, relations between research practice and care for the research participants and objects of inquiry are addressed rather rarely (see Lury, 2020 for extensive discussion of methods performativity; Sandberg et al., 2022 for discussion about caring for data in digital migration studies).

Some scholars explicitly reflect on the interplay between research methods and ethics (e.g. Nind et al., 2013), research practices and the process of 'becoming a researcher' (e.g. Hultin, 2019) and other, invisible, processes within research (e.g. Garforth, 2012), and the role of research tools in academic 'thinking' and knowledge production (e.g. Konopásek, 2008; van Es et al., 2018, 2021). For example, Nind et al. (2013) discuss the relation between ethics, methodological innovation, and reflexivity, questioning whether methodological innovation is inherently 'good' in the beginning of

their analysis, the authors highlight how methodological innovation has to keep up with and balance the promises of integrity, respect to the study participants, transparency, justice, empowerment, democratisation of research, and quality. While it is not the goal of my thesis to discuss the datafication of academic research practices and knowledge production, the brief discussion presented here aims to illustrate the intertwined relations between research methods, the researcher, the studied empirical processes and phenomena, and different kinds of knowledges. Attending to the methods handbooks and textbooks as foundational literature for building methodological expertise; I, first point to the need of further methodological reflection both on digital research and, specifically, on data studies if these are to become an established field of research. Second, I briefly discussed the ways in which research methods broadly, and for studying datafied societies specifically, are described and reflected on. This brief overview provides a springboard for further discussion about methods performativity (Barad, 2007; Law, 2004; Mol, 2002), ontological politics and politics of method (Fourcade, 2007; Mol, 2002), and their relation to the kinds of knowledges about datafication processes produced in empirical research.

3.3 Critical discourse on the social lives of methods assemblages

The academic discourse on the social life of methods[6] set out after a seminal paper by Mike Savage and Roger Burrows (2007) in which they described "the coming crisis of empirical sociology". In that paper the authors argue that methods such as in-depth interview and survey are dated and will not stand to all the needs of digital sociology. Even though Savage and Burrows did not attribute the coming sociological crisis solely to the problems of (digital) methodology, but also to the use and ways of dissemination of research information (2007, p. 887), the debate following the paper focused primarily on research methods. The academic discourse was further facilitated by a research theme "The social life of methods", co-convened by Evelyn Ruppert, John Law, and Mike Savage within the Centre for Research on Socio-Cultural Change (CRESC). The theme framed extensively a number of publications and editorials in the early 2010s. With the framework of the 'social life', Law, Ruppert and Savage (2011) critically engage with the representationalist approaches to methods and introduce arguments for the methods' sociality. Thus, the discussions of methodological innovation and methods attend not only to the ways of producing new kinds of insights with value of commensuration or conveying the 'proper' techniques for data collection, analysis, and presentation, but also as devices actively shaping empirical research processes and knowledge production.

Scholars participating in the social life of methods discourse challenge representational view on methodologies. Law and colleagues (2011) argue, first, that research methods are a part of and configured by a social world, and therefore are *of* the social. Therefore, research methods do not solely represent the (social) world, but are actively configured by it, not at least during the research process. Second, methods also configure the social world and therefore *are performing* the social. So, Law et al. (2011) emphasise the recursive relationship between the "social" and "research" practices and reject the ambivalence between the culture and science, enacted through different methods. Being at the same time *of* the social and *performative to* the social results in a "double" social life of methods. In order to understand how this social life is shaping research practices and academic knowledge production, the authors invite us to think of what methods actually do. They conceive methods' doings as "self-fulfilling assumptions about the character of the world" (Law et al., 2011, pp. 11–12) building upon the concept of methods' performativity. Annemarie Mol (2002) beautifully addresses the concept of methods performativity, arguing that in

[6] While the term 'methods assemblage' is a central analytical concept for my thesis and is coherently used in the next chapters, in this section 3.3 I use the term 'method' following the argumentation and vocabulary of the literature reviewed here. In my reading of the academic literature on the double social life of methods, terms 'method' and 'methods assemblage' can be used interchangeably.

academic research, methods can be considered as mediators between an object of inquiry and the representations, knowledges about it researchers produce.

> "Methods are not a way of opening a window on the world, but a way of interfering with it. They act, they mediate between an object and its representations" (p. 155).

According to Mol, the question of method is not what they are, but rather how do they interfere with the (empirical) world and how to think and to communicate their interference. Hereby research methods are not seen as particular techniques or tools of information retraction or analysis. Instead, methods act as researchers' ontological and epistemological assumptions about the world, enacted in these techniques. In datafication scholarship, multiple authors are concerned with the challenges that empirical datafication research faces, particularly when applying digital, computational methods (e.g. Lindgren, 2020; Rieder & Röhle, 2017), for example in relation to what and how can be known when studying datafication processes and applying digital methods. While some authors argue for expanding academic competences of social scientists to the theoretical understanding of the concepts and techniques underlying computational methods (e.g. Lindgren, 2019, 2020), others propose to bear with the unknown and pay specific attention to what is othered, excluded, and invisible (e.g. Coleman et al., 2019; Lury, 2020; Piattoeva & Saari, 2020). Similarly to what Mol (2002) addresses as ontological politics, Law and colleagues (2011) also draw attention to the institutional context in that research is conducted and its relation to empirical practice of research, including methods.

Taken together, these four elements of the method—the researchers, their practices, the realities, and the context altogether can be conceived as *methods assemblage*. These elements of methods assemblages make it especially difficult to create and enact new assemblages and new practices, despite the technological changes evoked by digitisation and later datafication. Thus, in the core of the double social life of methods lies an appeal to consider research methods together with existing realities and knowledges in order to create new ones.

Following up these arguments, two special issues appear in 2013, which attend to methods as devices (Law & Ruppert, 2013; Savage, 2013). Law and Ruppert (2013) suggest an extensive metaphor of devices to illustrate the ways in which methods are set together and apart as assemblages. Devices are to be understood as heterogeneous arrangements set together through tinkering, which may include some elements and exclude others. Similarly, methods assemblages as performative devices include particular kinds of assumptions about the world, particular realities and knowledges, while setting others aside. According to Law and Ruppert (2013, pp. 232–233) only a part of these boundary-drawing processes is empirical, whereas the other part is in 'us' as researchers and our agendas. The latter notion particularly underscores the performative role of researchers. For the purposes of my argument, I note "*researchers*" as a category helpful to think with about the kinds of knowledges various research methods produce. Another crucial aspect noted by Law and Ruppert concerns the methods assemblages as boundary-making practices, demarcating the research object and defining what is part of the research process and what is not as the study goes on. While I attend to both concepts of methods' performativity and methods assemblage in the following sections, these approaches help to grasp the pressing questions of methods' role in the studies of the digital and the datafied.

Following Savage's (2013) reflections on why questions of method have become central to the academic debate, we turn our attention to the ways in which the social life of methods changes in the contexts of digitization and datafication. Savage frames it—fitting into the context of datafication—as

> "pick[ing] up from the recognition that standardized data today, especially that facilitated by digitalization, allows it to take on a lively form which exceeds the straitjacket imposed by positivist statistical procedures" (p. 6-7).

The challenge in reflecting on the social life of digital methods is, then, in understanding the kinds of assumptions about the world made through methods operating with digital data.

Ruppert et al. (2013) provide an overview over some of these challenges comparing digital methods as devices to the 1980s STS studies in laboratories. In contrast to laboratories, 1) digital devices and related methods are spatially dispersed and not concentrated in one place; 2) social science methods are (more) deeply than laboratories involved in the configuration of social life and humans; 3) digital methods rely on the data produced for other purposes than research in multiple ways. These challenges are further developed by Burrows and Savage (2014) as they revise their initial publication on the coming crisis of the sociology. First, the authors sketch the change from prevailing studies on discursive accounts of practices (retracted specifically for research purposes) to the studies on traces of these practices (often generated for other purposes than research). Continued in datafication research and related fields, attention to the studies of data traces has been criticised, as multiple scholars point out other kinds of data than inter- and transactional (see e.g. Beer & Burrows, 2013; Zuboff, 2015). Second, Burrows and Savage (2014) refer to the redistribution of methods introduced by Marres (2012): a displacement of research techniques between academic researchers and private corporations, the researched, digital environments, platforms and other human and more-than-human actors. Third, Burrows and Savage (2014, p. 4) identify a "self-referential performativity" in which specific patterns are being reproduced by a specific and highly visible group of actors, while other actors remain invisible. That reminds us how data *are social and performative of the* social and, as Uprichard (2013) critically mentions, at the same time are often used and conceptualised in a positivist tradition of thought. Particularly data studies building on a critical, reflexive approach to research methodologies oppose this positivist application of methods. As Whittaker (2021) argues, however, even critical scholarship sometimes fails to avoid the big tech industry's narratives and assumptions about the world and individuals inscribed in datasets and techniques that researchers repurpose.

Furthermore, Burrows and Savage (2014) highlight the changes in temporal structure of digital research. Applying digital methods allows to process and analyse big(ger) amounts of research data in shorter time than it was possible before, producing 'immediate' results and data visualisations, arguably ready to present to the publics. To illustrate the challenge, the authors provide a rather negative example of preliminary research results 'meddling' with the still ongoing survey (Burrows & Savage, 2014, p. 4). However, that example implies that if presented at a 'correct' point of time with the help of proper visualisations, digital methods are only to a small extent interfering with the empirical site in other ways, not at least due to their spatial distribution and distance to the researched site. That implication not only contradicts the challenge of methods redistribution, but also reifies the positivist understanding of data as detached from the empirical context. In datafication research, concepts like data and platform politics (e.g. Gillespie, 2010; Puschmann & Burgess, 2014) and affordances (e.g. Weltevrede & Borra, 2016) provide analytical frameworks to engage with the implications of the self-referentiality challenge. Data politics and affordances draw attention to data as being *of* the social. With the question 'when are data' Borgman (2015) similarly underscores the relation between digital data and the ways they are put to use both by practitioners and by researchers in their empirical work.

Further research currently addressing the second challenge—self-representation and reification of biases in digital data and data analysis tools—focuses, among others, on the algorithmic bias and oppression (Benjamin, 2019; D'Ignazio & Klein, 2020; Eubanks, 2017; Noble, 2018), decolonial data studies, data activism, especially in the Global South (e.g. Birhane, 2020; Milan & Treré, 2019), and data justice (Dencik et al., 2019). Specific to these academic fields is their common concern with political and societal implications of datafication processes and the role of researchers-*activists* in making inequalities and injustices visible and putting data to use for the communities. Scholars from these fields turn to a variety of methodological approaches, including qualitative and ethnographic research, action and participatory research or computational social science in order to unpack the 'self-referential performativity' of international corporations in the reproduction of particular social injustices through data. Activist and participatory research

approaches, setting out to *intervene* and purposefully configure the empirical field together with participants on site, become widely applied across various fields (see e.g. Couldry et al., 2015 for action research project on narrations; Jarke, 2021 on co-creation with older adults). Both action research and participatory research projects aim at giving a voice to the invisible, marginalised, and oppressed communities and actors and conducting research not *about* the people in an empirical field, but together *with* them (Costanza-Chock, 2020). Such research projects highlight in a productive way how engagement with methods assemblages for studying datafication processes unpacks biases in data and technology and makes the latter visible for the researchers and the related communities.

Another challenge particularly distinct due to the digitisation and following it datafication is the redistribution of methods. While Burrows and Savage (2014) discuss that as an example of how social research is intertwined with media and technologies, with the same notion Marres (2012) shows a way to engage with practices within the research processes and not outside of these. In the latter sense, reflecting on methods assemblages opens spaces for examination of the background and evolution of a particular data collection or analysis technique and the market value of data produced through these. In the former sense, methods do not only build bridges to the socio-technical elements of the researched empirical site, but also include their own material and embodied properties. We currently notice such engagement with the redistribution of research within the methods assemblage in the discourses on the materiality of research and methods, particularly articulated in the fields of inventive methods (Lury & Wakeford, 2012), walking methods (Jarke, 2019; van Es & de Lange, 2020), and methods attending to data movement such as data journeys (Bates et al., 2016; Medina Perea, 2021), and creative methods (Kara, 2020). Most recently, the socio-material aspects of traditionally computer science methods such as prototyping experienced a new wave of attention in the social sciences and humanities (Estalella, 2016; Lupton & Watson, 2020). However, the role of material artifacts and the materiality of prototypes have already been discussed, for example in the works of the anthropologist and STS scholar Lucy Suchman and others much earlier (Suchman et al., 2002). What all these approaches have in common is an understanding of methods as an assemblage distributed in space and time among multiple *human actors, technological infrastructures, technology, and technology providers*. I will return to these analytical categories in the concluding discussion of this section.

A number of recent publications engage with multiple ways to acknowledge and employ the social life of methods in empirical examinations of datafication and data practices. So, elaborating on digital methods Kennedy et al. (2015) turn to action research in order to investigate how practitioners in the UK public sector engage with the social life of methods. Coleman et al. (2019) notice the opportunities of a practice-based (or practice-led or practice as research) approaches that provide ways to think and enact research methods in new ways. Lury and others (2020) also attend to methods as practices and "means to build interdisciplinarity" (p. 33) engaging researchers in the re-negotiation of the social (see Ruppert et al., 2013) and the empirical that is not distinct from the research process but tightly intertwined with it.

From the methodological point of view, the latter argument illustrates how different elements of the research site, including digital (research) data, are performative of the kinds of knowledges produced through digital methods. Therefore, for my further analysis of methods and methodologies of datafication research, I borrow categories that are often not made explicit in written methodological elaborations, as Law and Ruppert (2013, p. 234) illustrate with their analysis of surveys. These categories include "the character of the social" and its elements: an ontological question of how researchers conceive *society* and what "society" includes, and the question of what, *whom, whose voices and whose problems* do the research methods and data represent as means for which *(research) ends*.

> "Here, questions of formal method can be associated with 'making explicit' what might otherwise be implicit. However, it should not be thought that this process of making explicit somehow eradicates the implicit – it simply displaces it" (Savage, 2013, p. 17).

Candea (2013) also examines the concept of fieldsite from the perspective of a performative methodological device and thus adds *spaces and places* to the analytical categories required to think about research methods and their role in academic knowledge production about datafication.

Despite the understanding of methods performativity, negotiated in the social life of methods debate, it invokes an epistemological perspective in which knowledge can be a result of cognitive, communicative (discursive, textual) practices and leaves out other ways of sense-making. Thus, despite the critique on the representational methods, the (double) social lives of digital methods bracket out the richness of some non-representational knowledges. However, as Kennedy et al. (2015, p. 175) notice, "social scientists already understand the double social life of digital methods" while they develop new perspectives to reflect on and "live" with methods' social lives in their empirical investigations. So, Back and Puwar (2012) created a manifesto for live methods, arguing for more tools of real-time investigation that should, among other aspects, attend to *historical* frames in *creative* ways and establish *emotional* relationships in the course of research. Creative and affective methods (Vannini, 2015) enact strong critique on the representationalist methodologies and produce empirically non-representational kinds of knowledges. As Coleman and colleagues (2019) put it together, non-representational methodologies enable us to understand that "meaning may not always be what is at stake in research as sensation, feeling, embodiment, interactivity and engagement may be generated, grasped, understood and intervened in" (n.p.).

Further endeavours to engage with methods as an object of inquiry in social sciences include methodography (Greiffenhagen et al., 2011) and frameworks embedded in the concept of the third space and in-betweenness (e.g. Hurdley & Dicks, 2011). Especially in interdisciplinary research settings, Woodward (2015) argues methods and research artifacts are displaced from their traditional disciplinary boundaries and open up a third space, where methods are openly negotiated. Through the dialogue between the methods and explicit-making of the gaps which particular methods leave in the empirical examination, new spaces for methodological and object-specific investigation open up. The dialogue between methods (especially when these are embedded in different theoretical or disciplinary grounds) almost necessarily produces some "displacement, detachment and disorder" (Hurdley & Dicks, 2011, p. 289), but also "closeness, empathy and [] knowing" required to produce a situated view on the object of study instead of the gaze from nowhere (see also Haraway, 1988). Thus, the 'messiness' of (interdisciplinary, multi-theoretical) research practices pave the way to the renegotiation of the social research envisioned by the scholars who facilitated the discussion of the double social life of method.

In sum, what follows from the critical discussions of the double social lives of methods assemblages in data studies are a number of research practices and analytical categories that foreground some of methodological challenges identified in the debate. As outlined in the section, these categories include:

- the researchers with their embodied emotional relationships within the empirical site, their ontologies, epistemologies, and personal or institutional goals in the broader frame of institutional /disciplinary research politics;
- the researched (persons and things) as well as their interests, problems, histories, and values;
- times, spaces, and technological infrastructures invoked at the empirical site and in which the research is distributed;
- research process, practices, and techniques used for academic knowledge production.

While this list is not exhaustive, it provides an overview over the complexity and richness of what can be included in a methods assemblage empirically. In the following chapters presenting my

empirical investigation, where research methods are the object of inquiry, I return to these categories as a means to define the methods assemblage.

The social life of methods perspective teaches us that in order to engage with methods as an object of inquiry, two elements are crucial. The first one is an understanding of methods as a performative assemblage, highlighting the recursive relationships between methods, data, the multiple persons, realities, and knowledges entangled in a single academic intervention (Estalella, 2016; Law, 2004). The second element emerging from the critical discourse on the social lives of methods is an acknowledgement of processual character of methods assemblages. Methods assemblages stretch in time and space beyond actual activities of data collection and analysis, and include practices sometimes conducted far beyond the point when a research project has started or the funding for it has been secured. Data studies are sensitive to the arguments put forward by the discussions about double social lives of methods as these arguments similarly apply to the understanding of data and datafication processes, as I have shown in the previous chapter. At the same time, while data studies are reflective about the relation of digital data, society, and knowledge, data studies still require further conceptual and empirical reflection on the place of methods assemblages in these relations. The aim of my thesis is to provide a heuristic for such a reflection. For that, in this section 3.3 I reviewed main arguments about the double social lives of methods assemblages with particular focus on data studies. In the next section, I elaborate in more detail on the concept of methods performativity and the role it plays for reflecting on methods assemblages in data studies.

3.4 Performativity and methods assemblages

Through the historical and critical overview presented in two previous sections, an understanding of methods as an assemblage crystalised, that includes a multitude of elements beyond the researchers, their techniques, and objects of study. The methods assemblage as a concept is grounded in the notion of ontological multiplicity, as developed by Mol (2002). As an organisational ethnographer in a hospital, Mol elaborates on 'multiple bodies' lived and enacted through lower body atherosclerosis by patients, doctors, and medical laboratories workers. In her ethnographic examination, Mol illustrates how each of these actors produce different realities of the disease. Respectively, atherosclerosis is a pain in the leg, a particular medical index, or a specific state of the blood vessels. The emphasis therein is on the ontological differences between the multiple 'diseases', enacted at the same time by different assemblages and through different practices under the name of atherosclerosis. Mol's main argument is that the reality is not described but rather enacted through specific practices in the field in multiple ways. Applying Mol's framework of multiplicity to the methods assemblages for studying datafication suggests that each assemblage is composed of different layers, each of which provides a single narrative, translated, and rationalised according to the researchers positionings in the field and their research interests. In the relational understanding of methods, that translation work is crucial, as ontologically a methods assemblage is not a sum of distinctive elements including a clearly defined researcher, research situation, techniques, etc., but rather an enactment of practices that connect and relate the assemblage at any point in time. The empirical challenge, therefore, lies in the ways to disentangle these different layerings required to understand datafication and data practices. These layerings delineate between the elements of the assemblage that are "in-here" and "out-there" (similar to how Marres (2012) addresses methods redistribution within and outside research practices), between the research subject and object, the researcher and the researched.

Law (2004) builds upon the notion of multiplicity developing a concept of a methods assemblage. According to Law, multiple realities draw a 'messy' picture for academic research, since seemingly one and the same research object may be changing its names and even shapes. Additionally, from the methodological point of view, research objects also "move about" in the

field (Law, 2004, p. 78) depending on the positions of scholars and their topics. For Law, the focus, thus, shifts from the multiple realities themselves to the question of how research practices partake in the production of these multiple realities. He addresses an answer to that question with a methods assemblage. Drawing parallels to Mol's work, Law directs our attention to the contingencies of research as making some of the multiple realities of the field 'present', 'absent', or 'othered'.

Coherent with the argument of multiplicity, in his book "After method. Mess in social science research" Law (2004) provides various conceptualisations of a methods assemblage. First, a methods assemblage moves beyond the representational view and is embedded in the relational ontologies, including processes and contexts as the core elements of the assemblage (p. 67), and therefore the assemblage cannot be fully definite. Having said that, it does not mean that an infinite methods assemblage cannot produce objective, accountable scientific results. On the contrary, disentangling the layerings of the assemblage, which situate that assemblage in the context and exploring the relations that enable to situate the assemblage, helps to produce objectivity in research. This view resonates strongly with the production of situated knowledges (Haraway, 1988), discussed in relation to data studies in the previous chapter. Second, a methods assemblage puts forward the relations between statements and matter (their material properties) that are required in order to hold the assemblage together and situate it in the broader context. Both these arguments foreground the practices enacting different layerings of the methods assemblage and drawing boundaries between these different layers depending on the context. Therefore, the final conceptualisation of the methods assemblage, suggested by Law (2004) reflects the boundary-making practices and focuses on "the crafting and bundling of relations […] into three parts: (a) […] present, (b) whatever is […] manifest in its absence; and (c) […] absent but […] other because, […] it is not […] manifest" (p. 84).

For Law, what is present can be any kind of process, object, or representation that is enacted "in-here", that is explicitly included in the assemblage. The manifest absence can be described as the other side of the coin of presence: every time something is present, some other 'thing' is absent, e.g. some context, other relations, or processes. Otherness can appear in forms such as routine, insignificance, repression, and all the varieties of 'other' that cannot be accounted for in practice or in science. Law (2004) gives an example of criteria used in the practices of alcohol consume treatment, addressing the availability of alcohol for purchase as an example of a not manifest otherness: something that is not part of the treatment criteria. A methods assemblage is, then, a way of setting and enacting boundaries between different forms of presence and absence. What methods do is define how these boundaries are set. Sometimes these boundaries can be hard and 'steady' as for example is the case for positivist methods, aiming to distinguish clearly objects of inquiry from other objects. Sometimes, however, methods help to grasp the complexity and fluidity of realities. As already noted in the previous chapters and sections, datafication increases the complexity of research practices. In the empirical research of datafication, both data and methods are performative and enact different assemblages that are difficult to disentangle.

This understanding of methods assemblage is widely adopted in recent studies based on relational ontologies (Estalella, 2016) alongside with various similar concepts foregrounding the complex relations enacted in the research practices (e.g. methodological apparatus as in Barad, 2007; or research-assemblage as in Fox & Alldred, 2018). Altogether, these notions lean on the concept of performativity of methods, challenging the representational methodological perspective. Barad (2003) argues that the scepticism towards representationalism developed only when the studies of science turned to the specific practices and dynamics of scientific labour, rather than they were concerned with knowledge production. Barad refers to 'materialisation'/'materialized refiguration' by Butler and Haraway and to other scholars like Latour and Rouse who did not (only) use the term performativity but developed similar concepts. For Barad's account of (posthumanist) performativity, it is crucial to understand how exactly discursive and material practices produce

matter and (de)stabilise the boundaries of given categories (Barad, 2003, p. 808). The author also gives an overview over the research domains where performativity is a continuous part of the discourse: theatre studies, performance studies, literary studies (p.807-8). Asking how discursive and material, bodily practices are connected, Barad concludes that the issues of power relations and their materiality are central. Challenging the idea that power is solely social, it is therefore required to ask about the material consequences of power. For Barad, it is also an argument for interdisciplinary work, because the power relations will go missed if we focus on tracing them within discipline boarders.

Barad's posthumanist account on performativity builds on the notion that social and power relations are connected materially. Her account advocates causal relationship between specific practices embodied in material configurations and their material relations (which Barad calls phenomena). Barad introduces that causal relationship as intra-action. Contrary to *inter*action, where pre-existing interacting entities are the starting point, intra-action determine the boundaries and give meaning to the parts of complex relations as they develop. To explain how boundaries in the intra-actions are drawn, Barad introduces the notion of an "agential cut". Agential cuts draw the boundaries between the entities as they intra-act. Barad argues that applying the concept of performativity to research methods helps to reflect on the ways in which researchers make "cuts" through their empirical sites and data in order to make sense of them. Similar to the Cartesian cut, agential cuts enacted in intra-actions also determine the distinctions between subject and object, although not ontologically, but locally, within the relation. Through situating the agential cuts locally through the detailed accounts of practices that enacted these cuts, the objectivity of research is prevailed. Methodological objectivity for Barad "*means being accountable for marks on bodies, that is, specific materializations in their differential mattering*" (2007, p. 178, original emphasis), where bodies do not necessarily mean human fleshy bodies but various kinds of things, including the units of analysis and empirical sites.

Barad (2003, pp. 815–816) explicates it with the methods used in physics to examine the diffraction of light: depending on the measurement test, light is either determined as a wave or as a particle, whereas light without a measuring or capturing apparatus is indeterminate. Here, only a combination of different methods assemblages determines what kind of results about the nature of light are acquired. In another example the person operating a quantum microscope and the microscope itself are entangled in a unique way that enables people to "touch" atoms, which is only possible in this particular assemblage (Barad, 2007). Without the microscope, the physicist is unable to observe atoms directly, and the microscope requires an operator; together they produce a methodological apparatus in that no distinction between the researcher and the device can be made as only together can they perform the empirical task. With an ironic motto "Not simply intervene, enact the between" Barad and other scholars (Hultin, 2019; Prinsloo, 2019; Springgay & Truman, 2018; Taguchi, 2012) argue that in social sciences, similarly, the choice of methods and methods assemblages directly relates to kinds of knowledges they produce about the world.

As researchers we interact with, produce, and change the data and datafication processes that are our object of inquiry. The concept of performativity helps us to think of and grasp the interactions and changes we make, although it does not provide us with a formula for the "delta" before and after interaction. Performativity teaches us, however, to perform agential cuts (Barad, 2007) and draw borders between us as interested, knowing, and feeling observers with our observation tools (the methods assemblage) and the intra-action we study with that assemblage. When we apply methods to study various elements/aspects of datafication processes as objects of inquiry using data infrastructure and sometimes even our own digital doubles, it is difficult to distinguish between all of the above. It is especially so when the platform where the research is conducted and the method assemblage overlap, as it often happens e.g. in studies of social media data. Hence, to continue research work and to produce accountable and reproducible—and therefore rigorous results,—we need theoretical tools. The concept of performativity proposes

agential cuts as an analytical tool to situate the multiple elements of a methods assemblage. Although Mol, Law, Barad and colleagues provide an outstanding analysis of their respective empirical cases, methodologically they do not equip us with explicit analytical and reflexive tools required for attending to the role of the methods in studying datafication processes. My thesis attempts to extend the performative approach of methods assemblage onto the analysis of datafication and develop a heuristic that provides datafication and data scholars with vocabulary sensitive to the multiplicities and complexities of the entanglements between research processes, data, society, and knowledge.

3.5 Assembling carefully

In the beginning of this chapter, I briefly revisited some of the main arguments of the sociology of scientific knowledge production. While these studies to a great extent elucidate how knowledge is produced in natural sciences and their laboratories, using the notion of 'dissecting room' discussed by Knorr-Cetina (2002), I argue that empirical datafication research, similarly is detaching (at least partially) some aspects of the datafication processes. In this process, new conceptualisations of 'datafication', analytical categories, and definitions are being developed. Knorr-Cetina (2002) argues that through new classifications/definitions, the role of the analysed objects of study and their functions within a certain social reality are established in relation to its other aspects (p. 165). While datafication research creates some sort of a dissecting room, at the same time, it also participates in datafication processes (e.g. when using computational techniques of data collection and analysis, but also when engaging with and rendering visible otherwise hidden aspects of datafication). With that, conceptualisations of datafication, produced by empirical scholars, when produced, subsequently re-situate these datafication processes in the broader societal reality.

As I demonstrate in the next chapter, my thesis aims to explore how this detaching and re-situating works. For that, I turn to methods assemblages as practices required for this detaching and re-situation: by enacting a methods assemblage, various elements of empirical datafication phenomena and research practice are assembled that bring forth re-situated conceptualisations of datafication. To understand methods assemblages in social sciences, I also briefly elaborate on methodologies, epistemologies, ontologies, and their interrelation in regard to data studies and datafication scholarship. Historical developments of different qualitative and quantitative research methods illustrate how and why current datafication scholarship is primarily built on the qualitative research projects. A brief review of some methods handbooks and textbooks provides a more practical perspective on the role of methods in research processes, the kinds of knowledges, and research outcomes that can be produced with various methods for studying datafied societies. A critical and reflexive view on methods is presented through a discussion of the double social lives of methods. The arguments about methods' performativity, relationality, their recursive relation to the empirical research site, and academic knowledge production resonate with the arguments about relationality and recursivity of data that lie at the core of data studies. Many datafication scholars elaborate on the complex and complicated relation between digital data as research object, various methods for studying datafied societies, technology required both for research and empirical practice, and our knowledge about datafication. Introducing the concept of methods assemblages embedded in the understanding of methods performativity, I apply an analytical tool that should help to describe and understand these relations better.

The questions like "Cui bono?" posed by Star (1995) in the quote in the beginning of this chapter come from critical and feminist traditions of thought, asking about the work and the power 'behind' academic knowledge production and how these are distributed among different stakeholders. As the previous chapter 2 has shown, data studies are concerned with similar questions in relation to datafication processes. This chapter illustrates how social researchers examining science and academic knowledge production also need to *care* about these issues. So,

some scholars have adopted a notion of care to their reflection on STS research, arguing that care-ful research "reflects the iterative development of appropriate concerns, sensibilities and research questions" (Law & Lin, 2020, n.p.). In another piece, Law (2021) further develops the notion of care-ful research, addressing how careful, close reading, slow research are "holding open differences and awkwardness and tensions within research" (p. xix). Making this explicit in thinking and writing allow new, different, productive kinds of critique. Other scholars specifically propose so-called 'slow scholarship' opposing neo-liberal practices in academia (although this discussion is not part of my research), encompassing such practices as collective writing and caring for oneself (e.g. Mountz et al., 2015). Feminist concept and ethics of care are helpful here as they not only engage with the ways to produce situated knowledge, but also consider ethical and political dimensions of knowledge production. They conceive of care as ethico-political obligation (Puig de la Bellacasa, 2017), a moral 'disposition' (Tronto, 1993), and a local, situated practice (Mol, 2008) that are sensitive to otherness, affect, pain, and trouble and are directed at creating better worlds. Overall, research related to the concept of care can be understood through two different lenses: 1) a normative, ethical concept advanced by e.g. Tronto (Tronto, 1993, 2016) and Puig de la Bellacasa (2011, 2017) and 2) an ANT-inflected perspective on care, for example developed by Annemarie Mol and colleagues (Mol, 2008; Mol et al., 2010) that foregrounds the local, situated practices in which care is performed. Recently, increasing body of research aims to combine both normative and practice-oriented perspectives on care (e.g. Lindén & Lydahl, 2021) in order to produce more generative, inclusive views and conceptions about the social/socio-technical realities. Particularly when rapidly developing datafication processes complicate academic knowledge production and datafication scholars are negotiating the limits of what can be known about datafication processes, the techniques and tools required for that, and the power of certain actors to decide who gets access to which digital data, a reflexive and care-ful approach to assembling research can be helpful. To capture performativity of the method assemblages methodologically and empirically, in the following, I address research methods in their practical enactment. To be able to do so, in the next chapter I briefly discuss methods assemblages-in-practice as an analytical lens of my thesis before I proceed to the description and reflection on the research design of my project.

4 Research design

My dissertation follows the argument on the double social life of methods and methods performativity, presented in the previous chapter:

> "In a world where everything is performative, everything has consequences, there is, as Donna Haraway indicates, no innocence. And if this is right then two questions arise: what realities do the current methods of social science help to enact or erode? And what realities might they help to bring into being or strengthen?" (Law & Urry, 2004, p. 396)

In this chapter, I elaborate on my research design developed to study methods as an object of inquiry through the analytical lens of methods assemblages and reflect on the kinds of realities my research design enacts. For John Law, methods assemblages bring together various actors involved in academic research through practices of inquiry. He addresses this as "method-in-practice" (Law, 2004, p. 45). Following that argument, my research design extends on two practice-based conceptual approaches (Nicolini, 2009b; Schatzki, 2002) that allow to address the performativity of methods assemblages in practice. Applying practice-based concepts developed by Schatzki (2002) helps to understand methods assemblages as both "doings" and "sayings" (p. 72): empirical datafication research and its results reported in academic writing. An iterative approach of zooming in and zooming out developed by Nicolini (2009b), in turn, allows to consider the practices that comprise a methods assemblage and situate these in the broader context of various academic disciplines, related research politics, and partial realities of scholars conducting data inquiry. In my research design, I alternate between zooming in onto each of these practices and zooming out onto the academic fields and research politics to situate my findings. I am drawing on theories of practice as an analytical lens of my thesis that helps to explore methods assemblages not as a stable, finished relations between all elements of a research process, but also as an ongoing process that includes reflection of research conducted in the past and planned in the future.

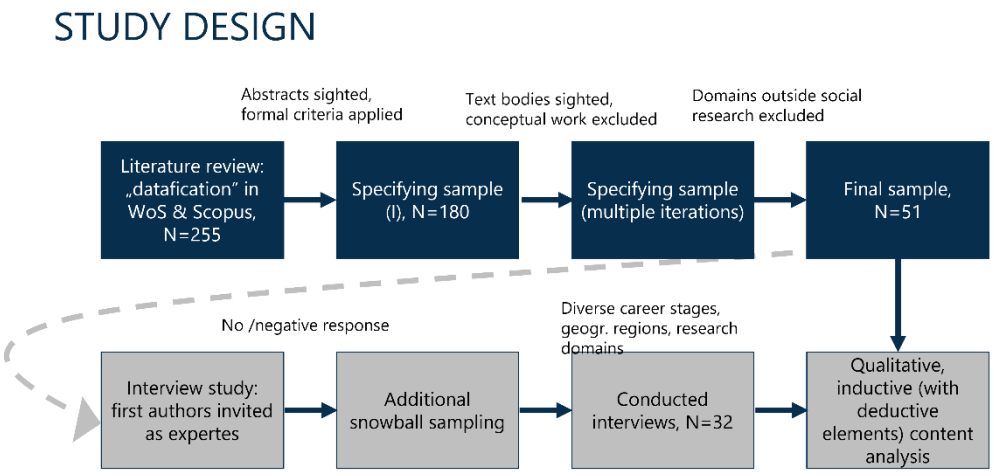

Figure 4-1 Research design overview

Figure 4-1 provides an overview over the research design. In order to map out what kinds of methods assemblages are applied in the current empirical research on datafication in social sciences and what kinds of concepts about datafication they produce, I developed an iterative research design applying a mix of quantitative and qualitative methods. I conducted a literature synthesis of datafication scholarship and expert interviews with authors of the sampled literature. A literature

analysis, including quantitative keywords analysis, zooms in on practices and elements that comprise each methods assemblage. The expert interviews study with the authors of the sampled publications provides a deeper insight into the practices of data inquiry and situates the findings of the literature analysis in the various contexts of academic scholarship. Expert interviews zoom out and situate methods assemblages as one of many parts of academic scholarship, alongside with research politics, disciplinary and domain-specific discourses, and researchers' own subjectivities and positionings towards datafication processes they study. Finally, sampled articles were analysed using a mix of deductive qualitative and quantitative analysis. Interviews were analysed qualitatively, while developed inductive categories were central to the analysis. Although the figure 4-1 and the description provided in this chapter follow a chronological line, in practice my study was conducted iteratively.

The iteratively developed literature synthesis maps out empirical datafication scholarship in social sciences published in 2013-2019. The literature sample was developed based on the keyword 'datafication' and its word forms. Only original full research articles published between 2013-2019 in social sciences and listed in the Scopus and Web of Science databases were considered. To address the methods assemblages in practice, according to Law (2004), I included only papers presenting empirical work. As I discuss later in this chapter, in contrast to conceptual contributions, empirical articles report about research projects conducted by datafication scholars at an empirical site of practice, while methods and said sites of practice are described explicitly. According to my research interest in the data inquiry in social sciences, I identified the keyword 'datafication' as a suitable search term for my empirical project. The term 'datafication' is currently widely used across academic domains and fields (Couldry & Hepp, 2017; Jarke & Breiter, 2019; Schäfer & van Es, 2017; Williamson, 2017), bundling together different theoretical, epistemological and ontological approaches to data inquiry. As a keyword, 'datafication' does not have further colloquial meanings, which would be the case for i.e. 'data', for which the meanings can reach from digital traces to empirical research material. The timeframe for the literature synthesis is chosen according to the historical development of the term 'datafication' in social sciences that begins around the year 2013 (see fig. 4-2) with the publication of the book "Big data: a revolution that will transform how we live, work and think" by Mayer-Schöneberger and Cukier (2013). An initial keyword search for 'datafication' in the title, abstract or keywords in the academic databases Scopus (https://www.scopus.com) and Web of Science (WoS, https://www.webofscience.com) confirms that the term has been used widely only since 2013. Figure 4-2 illustrates the result of this search in Scopus and shows a significant increase in the number of contributions with the keyword 'datafication' from the year 2013 onward. An analysis of publications addressing datafication in the past (nearly) ten years, therefore provides a comprehensive synthesis of empirical scholarship from various domains of social research and is sensitive to its historical developments.

Both chosen scientific databases index curated content (e.g. academic articles, books, book chapters, conference proceedings) from publishers selected by a review board/expert team of each of the databases. Such indexing allows databases' users to search through content published by a variety of publishers and in many journals, rather than search each publishers' or journals' archives. Both databases, however, require licenses for accessing them, which in my case were provided by the University of Bremen. Since each database has their own selection procedures and guidelines, their indexing differs, and search results of the same search query might differ from one another. For this reason, both databases were used to sample the literature for my synthesis. Both databases not only provide tools for searching academic literature, but also multiple metadata relevant for literature analysis such as information about the authors, about publishers and outlets, assigned keywords, references, and citations, among others. In contrast to Scopus, WoS consists of several literature search databases, organised around different aspects such as subject focus, document types, datasets, or certain geographical regions (see https://clarivate.libguides.com/authors). Searching both databases can be organised around disciplines, the attribution of academic content

to certain disciplines and the granularity of disciplinary fields available for search slightly differ between the two databases. Combining search queries from both databases, therefore, ensures that more content is included in the initial sample.

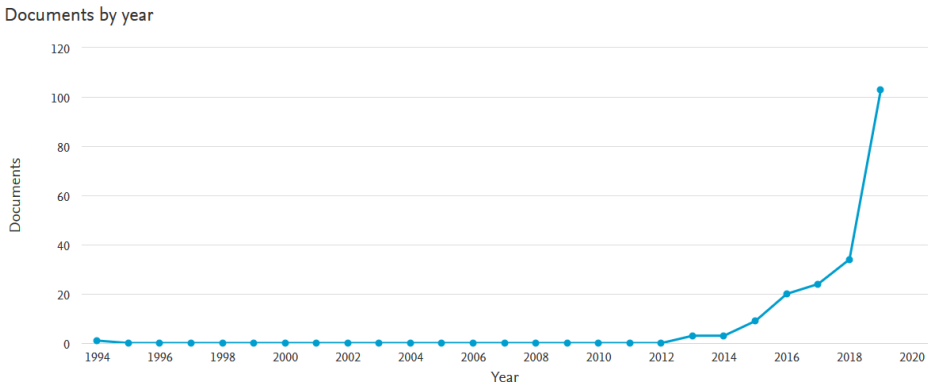

Figure 4-2 Scopus analysis of search results for the keyword "datafication" in the title, abstract or keywords of research articles in the source type journals in the domain of social science until the year 2019, without limitations to the language of publications. Visualisation created with Scopus.

This chapter begins with a brief overview of practise-based theoretical and methodological approaches that provide conceptual background of my research design. In the next section, I present in more detail design of my literature synthesis and reflect on quantitative and qualitative methods of literature analysis. Next, in the third section, I elaborate on the interview study with datafication scholars and reflect on conducting virtual interviews during the pandemic. Summing up my research design, in the fourth section of this chapter, I briefly address limitations and challenges of my project and elaborate on the methods sensitivity pursued in my study in contrast to other, methods-centric approaches. Finally, as the whole world was struck by the outbreak of the Covid-19 pandemic in December 2019-early 2020, I reflect on its impact on the social lives of my own research project and the role of datafication processes in my doctoral journey in the concluding section of this chapter.

4.1 Practice theories as research design

"The inquiry needs to be practical: an exploration of method-in-practice" (Law, 2004, p. 45). With this line in his book "After method: mess in social research" John Law not only calls for more empirical research on methods, but also provides a concise summary of the concept of methods assemblages developed in this book. A methods assemblage, thus, can be understood as a bundle of practices ordering and associating some human and non-human actors situated in certain times and spaces while excluding others. In my methodological exploration of current datafication scholarship, attending to practices empirically and analytically allows to achieve two goals. First, *a focus on methods assemblages as practices sheds light on how different elements of the assemblage are drawn together and how they produce certain knowledges about datafication processes.* Understanding practices comprising methods assemblages paves a way to exploring the performativity of methods assemblages empirically and furthers the theoretical discussion on the double social life of methods presented in the previous chapter. Second, *assuming practices as a central analytical concept allows navigating interdisciplinary research designs applied in datafication scholarship to multiple, heterogenous empirical cases.* For my research project within and about data studies presented here, practice-based approaches provide an analytical lens for studying empirical research grounded in diverging theoretical, epistemological, and empirical domains alongside each other. Following Nicolini's (2009b, p. 1393) call for eclectic theories and methodologies in practice-based studies, I weave together theoretical

concepts of methods performativity, methods assemblages, and practices in a multi-method study of datafication scholarship. I draw on theories of practices as an analytical strategy for my empirical study applying which I developed my research design. In this section, I briefly review practice-based theoretical and methodological approaches relevant for my project and discuss the advantages and challenges of applying these to the studies on research methods and datafication processes.

4.1.1 Understanding practices

What is often described as the practice theory is a set of different ontologies and epistemologies, which demarcated a 'practice turn' in social sciences and STS. The development of what is usually addressed as practice theories began in the 1960-70s. As an alternative to positivist ontologies, theories of practice aim to overcome the Cartesian challenges of a researcher as a third-party observer, and methodological, epistemological individualism (Schäfer, 2017, p. 36; see also Gentzel, 2019) and "see themselves as situated knowledge that relinquishes the classic, universal claims of validity" as Jonas and colleagues (2017, p. 252) write referring to Hirschauer (2008). Within STS and social sciences, different concepts of practices have been developed (Hui et al., 2016; Nicolini, 2017; Reckwitz, 2003; Schatzki, 2002; Shove et al., 2012). As Reckwitz (2003) discusses, in STS, theories of practice were widely used by sociologists of science, allowing to attend to sciences as widely informal work practices, tacit—implicit—knowledge of individual actors, and the use of various non-human artifacts as part of practices of science (p.284-285). For Schatzki (Hui et al., 2016; Schatzki, 2002, 1996), practices are conceptualised as situated doings and sayings, which 'bundle' into socially-constructed practices when the actors recognize and perform their activities as such. Subjects, for Schatzki, are primarily humans as they are able to understand practices as such and reproduce them according to their rules and teleo-affectivities—"set[s] of ends, projects, and affectivities" (Schatzki, 2002, p. 80). Non-humans, artifacts, and things are acknowledged as relevant aspects of enacting practices. Practices are performed/enacted by actors at the sites of practice "where entities are intrinsically part of their own context" (Schatzki, 2002, p. 65). The site of the practice is a broader context, in which a practice is situated and performed, involving various related actors and discourses they are engaged in. Nicolini (2017), in turn, defines practices as regimes of performance—while performances are the empirically observed iterations of practices—and criticises Schatzki for defining practices as "some-thing" (p. 21). Nicolini draws attention to the multiple ways in which different practices relate to each other such as conflicts, interferences, or co-evolution (ibid., p.30).

Within media and communication studies, Couldry (2004) builds his practice-theoretical approach for media and communication studies on Schatzki's concept of practices and foregrounds questions about "what types of things do people do in relation to media? And what types of things do people say in relation to media?" (p.121) in order to move away from predominantly discursive studies of media. Further, Couldry (2004, p. 127, see also 2012) points to a pressing question of "potential hierarchies between media practices and other sorts of practice" (p. 127) that echoes concerns regarding performativity of digital and datafied methods discussed in the previous chapters (e.g. Lindgren, 2020; Rieder & Röhle, 2017). In media studies as well as in sociology practice-based approaches are also used for the study of the everyday (Rinkinen et al., 2015; Shove et al., 2012), media and the body (see Postill 2010), journalism (Witschge & Harbers, 2018), and media activism (Fotopoulou, 2019; Mattoni, 2017; Stephansen & Treré, 2019), among others. Especially in the domain of media and data activism, the notion of "acting on" media as a way of transformative, configuring activist work has been taken up by other scholars (Kubitschko & Kaun, 2016; Milan, 2019; Stephansen, 2019). Kubitschko (2018) applies the concept of acting on by extending practice-based research in media studies to the politics performed in practices and explicitly acknowledges non-humans (media, infrastructures) as a significant part of practices:

> "acting on denotes the efforts of a wide range of actors […] to take an active part in the moulding of the media technologies and infrastructures that have become part of the fabric of everyday life" (p. 633).

Despite the differences, various practice-based approaches share some commonalities, including flat ontologies, equal attention to discourses and material, physical, embodied activities, and understanding of practices as historical and durable (Pentzold, 2015).

Although the richness of insights offered by the theories of practice is widely accepted across research domains, they also face critique. The most frequent criticisms include 1) the inability of a flat ontology to illustrate the complex power interrelationships (Couldry, 2020), 2) the dissolving of an acting human subject in the practices and deprivation of cognitive and affective components of actions (Gentzel, 2015), 3) strong focus on routinised actions, neglecting innovative and creative work (see for discussion of this criticism e.g. Stephansen, 2019), 4) and a dominant attention to material, non-human artifacts and their agency within social practices (however the need to re-estimate the role of materiality in actions is acknowledged even among the critics, see e.g. Gentzel et al., 2019). In addition, within communication and media studies, the ability of practice theories to highlight global power relationships and macro-scale, historical events in general has also been questioned (Postill, 2010, p. 18; see also Gentzel, 2015). Critics argue that flat ontologies are not able to distinguish between different uses of power and different ways in which power can be enacted, therefore undermining the ability of social sciences to investigate power-ridden social relations.

4.1.2 Understanding methods assemblages through practices

In my thesis, I aim to overcome some of these challenges of practice-based approaches by combining the concepts of methods assemblages, methods performativity, and social practices in accordance with Nicolini's (2009b) "principle that treating practice in descriptive terms is often not enough and that a coherent analytical stance is necessary" (p. 1393). Such theoretical multiplicity provides a way to explicate which elements of the assemblage are to what extent involved in its ordering and how they allow to enact methods assemblages in different ways. In a published piece of a dialogue between Vicky Singleton and John Law (Law & Singleton, 2013, p. 486) Singleton notices that "knowing is embodied, situated, and embedded in practices, and practices are always being done somewhere." Some media scholars share such perspective on knowledge as "practical understanding of what is socially actionable" (Couldry, 2020, p. 1139; see also Addey & Piattoeva, 2021a; Law & Singleton, 2000). In this spirit, Stephansen (2019, p. 190) argues that

> "activities such as theorizing, reflecting, and analysing should themselves be treated as social practices, and that these kinds of 'knowledge practices' should be analysed as a core dimension of media practices."

Expanding on this notion and on the practice-based approaches discussed in this section, I propose to address knowing and producing conceptualisations of datafication processes—bundles of social practices performed by various heterogenous actors located at empirical sites of practice—as a constitutive element of the methods assemblages. An empirical site of practice, then, can be defined as spaces and times where and when methodological interventions and practices of academic research and knowledge production are performed. This understanding of methods assemblages is specific to an exploration of empirical social research—an application of methods-in-practice. Empirical research, following the argument made by Ruppert (2013) in the wake of 'big data' challenges for the social sciences, will be understood here as doing "immersive […] data work by innovatively, critically and reflexively engaging with new forms of data" (p. 270), while these new forms of data are generated, processed, and used by different stakeholders. Empirical research 'doing immersive data work' engages with social realities and relations of these various stakeholders and investigates the role of data and datafication processes in them.

If we consider practices as described by Schatzki and Nicolini in terms of methods assemblages, the chosen practice-theoretical approaches open interesting perspectives on the role of methods assemblages in the research processes. First, methods assemblages—including not only human but also non-human elements—can be placed in the domain of social orders, according to Schatzki (2002). Taken together with Nicolini's (2009b) notion that practices are institutionalised, it can be concluded that methodological practices control and enact power relations and acquire normativity in academic knowledge production. This conclusion is similar to Barad's (2007) interrogation of methods as drawing boundaries between the researcher and the researched in their intra-action. However, the idea of methods assemblages as institutionalised and normative orderings bears further consequences. First, normativity of particular elements of methods assemblages (e.g. within disciplinary settings) explains the challenge identified by Law et al. (2011) that the social life of methods makes it particularly difficult for researchers to produce radically new kinds of assemblages and knowledges even if the empirical site is experiencing technological transformation. Second, as normative orderings, methods assemblages also actively engage in shaping knowledges as they are othering knowledge that does not fit the given methodological 'order'. Thus, applying the categories of present, manifest absent, and othered, as developed by Mol (2002) and Law (2004) to the academic knowledge production about datafication processes through methods assemblages, unpacks what knowledges and respective conceptualisations of datafication processes are either methodologically 'normalised' or excluded. From that point of view, the accomplishments of methods-assemblages-in-practice, such as e.g. published academic articles or books and reports, also contribute to the reification of methodological 'order'. In hindsight of the discussions about the digital, digitised, and computational techniques often applied in social studies of datafication processes, an attention to methods assemblages as ordering academic knowledge production provides leeway for questioning and critiquing, how these orderings come to be and whose teleo-affectivities are guiding these orderings. Attending to methods assemblages in data studies from this perspective allows to critically engage with these questions already during the empirical research. Finally, understanding methods assemblages as including institutionalised and historically durable practices renders visible that some of their elements are not under discussion, at least within the frame of their respective disciplinary or theoretical settings.

4.1.3 Navigating between heterogenous practices empirically

With this understanding of methods assemblages and sites of practice, I apply practice-based methodologies in my study. Practice-based methodologies take the core characteristics of practices reviewed here as a starting point and are closely related to ethnographical or ethnomethodological approaches (Jonas et al., 2017, p. 254). Prevalence of observations over qualitative interviews as a method of data collection can be explained with "the observation that interview data ultimately only express ex-post-rationalisations ono the part of the interviewees about their practices (ibid, p.254-55). In addition, practice-based methodologies often call for self-reflexivity (Schmidt, 2017) and a multi-sited analysis of practices (Nicolini, 2017). To explore practices in the situations that cannot be directly observed or have happened in the past, scholars turn to the notions of translation and reconstruction (see Gherardi, 2019). For example, methods such as an interview with a double (Nicolini, 2009a) or shadowing (Czarniawska, 2007), ethnographies of an object or affective ethnographies (Gherardi, 2019) provide opportunities to reconstruct practices in detail. Similarly, scholars investigating the everyday practices also developed approaches to explicitly explain the practices performed with (everyday) objects. Rinkinen and colleagues (2015) apply diary-based research to explain how people encounter, act in, and evaluate the "object-world" in their everyday practices. In contrast to studies predominantly focused on the sense-making practices, Rinkinen et al. elaborate on how doing something with objects, performing evaluative acts (based on affect, personal values, etc.) also is part of socio-material practices attributed to human subjects.

For Nicolini (2009b), theories of practice are constitutively methodological. He proposes a methodology of zooming in and zooming out that requires theoretical and methodological multiplicity. The leading concern of the zooming in onto specific sites of practice is to apply multiple theories to both describe and understand practices. Zooming in on the practice means to study how a practice is accomplished, while zooming out of practice means to study how and with what effects is a practice associated with others across different sites. Zooming out to the associated practices helps to appreciate the 'horizon' of change and the effect of the practice under study. The methodological and theoretical multiplicity in social sciences can be achieved not only within practice-based approaches: triangulation (e.g. Flick, 2011) or bricolage (e.g. Denzin & Lincoln, 2018) have been widely explored and applied empirically since 1970es. All of these approaches aim at gaining a deeper and more detailed understanding of the studied empirical phenomena and practices from different perspectives. In my research design building on the methodology of zooming in and out (Nicolini, 2009b), I follow these principals of theoretical and methodological multiplicity. I apply a variety of concepts (methods performativity, methods assemblages, theories of practice) and methods of data collection and analysis (literature analysis, expert interviews, formal and qualitative content analyses, and quantitative keyword co-occurrence analysis) to understand what methods assemblages are currently applied in datafication scholarship and what concepts about datafication they produce.

In sum, I expand on practice-based approaches developed by Schatzki and Nicolini in my analysis, while I acknowledge the critique of theories and methodologies of practice discussed above. Situating methods assemblages in the empirical sites of practice in which datafication processes take place allows including a broad range of stakeholders and the implications they experience both from the datafication processes and research interventions. To overcome the common critique of practice-based approaches, I also attend to practices as inherently social, recognizing the agency of individual human actors—researchers and study participants. Following Schatzki (2002), human actors are cognizant and expressive. The practice-based approach chosen in this thesis serves to account as fully as possible for different elements of the methods assemblage while giving precedence to human agency. Further, my practice-based approach allows an analysis of heterogenous research designs, stemming from different disciplines and schools of thought. Understanding methods assemblages as practices means understanding them as ordering of various elements such as doings (empirical data generation, analysis), sayings (research documentation and reported results), that are situated in broader contexts (academic knowledge production, research politics) and provides tools for comparability of these heterogenous research designs. It allows a level of abstraction that moves beyond domain-specific and disciplinary differences, but maintains a situated, partial view on how methods assemblages are performed by whom, when, and why. Theories and methodologies of practice, in sum, provide an avenue to trace the entanglements of research practices such as research design, data collection, analysis, and reporting, with the empirical sites of practices where datafication processes are being performed.

Empirically, for my conceptual and methodological inquiry into data studies, understanding methods assemblages through practices is helpful in two ways. First, it allows understanding my literature analysis as an exploration of the accomplishments of practice (Nicolini, 2009b, 2017) that are not separated from practices themselves and/or "doings" (Schatzki, 2002), but rather present a continuation of these. By synthesising literature analysis with interviews with sampled articles' authors, I could not only gather accounts of the authors' previous research practices, but also enacted new ones. As Silvia Gherardi—a sociologist of organisations and work and a practice theorist—argues, "the reflection on how we do empirical research is an epistemological reflection about how 'things' are made to matter and how epistemological relations make 'things' acquire a situated position" (Gherardi, 2019, p. 2). Thus, attending to the expert interviews as practices of reflection is the second advantage of the practice-based analytical lens of my thesis. Through this reflection by and with the datafication scholars I became a part of the methods assemblage enacted

during the expert interviews. Focusing on methods assemblages as practices of ordering and assembling different elements of the methods assemblages for studying datafication processes helps to navigate within interdisciplinary research designs and various empirical sites of practice explored by datafication scholars. The following section demonstrates this empirically by outlining the details of my literature synthesis.

4.2 Literature analysis

To specify a literature review as a methodological approach, some authors (e.g. Onwuegbuzie et al., 2011) recommend the term research synthesis. The term 'synthesis' expands the scope of considered literature to include not only research articles, but also so-called 'grey literature' like working papers, reports, academic blogposts, websites, and other (e.g. visual) content (Onwuegbuzie et al., 2011, pp. 186–187). In my thesis, research synthesis covers a significant part of my study and I understand it as a methodology to analyse existing work on datafication. For my literature synthesis, I completed an analysis of academic publications with interviews with the authors of the sampled articles: together, I consider these my methodological synthesis. During the interviews, discussed in detail in the next section 4.3 of this chapter, interviewed scholars mentioned their various research projects, working papers, and reports. For the sake of anonymity of research participants these materials are not systematically included in my analysis, however I sighted them, which contributes to the results of my analysis. The Cambridge English Dictionary defines synthesis as "the mixing of different ideas, influences, or things to make whole that is different or new."[7] In this spirit, rather than extending my analysis with grey literature or systematic exploration of other, e.g. web materials related to research projects reported in sampled academic articles, I expanded the synthesis by conducting interviews that allow to situate studies on datafication reported in academic journals in the practices of conducting research and in application of methods-assemblages-in-practice. I conducted a literature synthesis of *empirical* work on datafication. The synthesis aimed at mapping out the field of empirical inquiry on datafication and applied methods assemblages.

As the word 'synthesis' implies, literature review does not solely aggregate knowledge, but provides a conceptualisation of the current research on a topic. A research synthesis is a challenging process, as it requires a cohesive understanding of not only research arguments in the analysed literature, but also the underlying ontological, epistemological, and theoretical concepts. Especially in the social sciences, the underlying assumptions are often not stated explicitly (Hart, 1998, pp. 11–12). Therefore, it is up to the researcher herself to recognise and identify ontological and epistemological assumptions of the authors. In my project, I further expand the literature synthesis and include personal accounts of the authors through the expert interviews. Especially when it comes to the analysis of research methods and methodologies, an openness toward different perspectives and a reflection of researcher's own theoretical situatedness are pivotal for a comprehensive and cohesive literature analysis. For a literature synthesis methodology to comply with the quality criteria of academic research, the methods ought to be systematic, explicit, comprehensive, and reproducible (Fink, 2005, p. 17). The self-reflection, hence, contributes to meeting these quality criteria.

The questions about what methods assemblages are applied in empirical studies of datafication and what concepts about datafication they produce is leading my analysis of academic publications. The goal of the literature analysis is therefore, to map out the variety of methods, fields, and topics related to datafication. To reach that goal, I iteratively developed a sample of research articles on datafication scholarship. I identified suitable keywords and scientific databases,

[7] https://dictionary.cambridge.org/dictionary/english/synthesis

developed inclusion and exclusion criteria for the sampling, and applied quantitative and qualitative (deductive, inductive) methods of analysis. Figure 4-1 illustrates the synthesis procedure.

Construction of a sample for literature synthesis is a trade-off between including a very broad variety of contributions and creating an overly specific sample, which does not allow for much variation in the analysis. After trying various systematic strategies (based on journals and publications metrics), I settled for a strategy based on keyword-based sampling. I used 'datafication' and further word forms such as 'datafi*' and 'datafy*' —addressed as 'datafication' in the following—to search Scopus and Web of Science databases in February 2020. The sample, therefore, included articles published until the year 2020, including online first versions of academic papers that were scheduled for publishing in 2020. I searched for 'datafication' in the title, abstract, and keywords in research articles in the source type journals in the domain of social science in English language in the Scopus database. I only searched academic articles published in journals; academic book chapters were not included for various reasons. Practically, it is less common for book chapters to enlist keywords; further, the duration from empirical study to the publication in an academic journal is usually shorter than the publication in an academic book: therefore, journal publications were chosen to include most recent research. The query returned N=165 contributions. These search results were used for the initial sampling. Considering that the Scopus database does not include all research on social sciences, an additional sample with the same search criteria has been drawn from the Web of Science database, which retrieved N=191 contributions. Both samples from two databases were then merged and duplicates were excluded, resulting in the N=213 original publications. After a proof-checking the search results two months later (April 2020), the same search (search queries were saved) retrieved N=196 results in the Web of Science database, the difference of five papers was identified as recently added to the database and added to the initial sample, resulting in N=218.

The main reason for additional search in April 2020, however, was to add academic articles using word forms of 'datafication' in addition to the noun. For that, initial search queries for both Scopus and Web of Science databases were modified. Instead of 'datafication', these included "datafy* OR datafi* AND NOT datafication" in the Scopus database and the query TS=(datafy* OR datafi* NOT datafication) in the Web of Science database. The search query with the Scopus database retrieved N=28 original articles. However, adding these further word forms to the query in the Web of Science database resulted in retrieving articles from different domains than social science: instead, many medical texts were identified in the latter query, which were not considered further. After adjusting the search query to include only social research domains, the search query in the Web of Science database retrieved N=32 original articles. After removing duplicates from this additional search, the total of N=42 academic articles were added to the initial sample. In total, after the additional search in April 2020 and further removal of duplicates found in the dataset, initial sample counted N=255 original papers (see Appendix 3 for a full list of references).

Subsequently, I analysed all titles and abstracts, and in the following step the whole text bodies, in order to narrow down the sample according to the developed inclusion and exclusion criteria. Only empirical work has been included, following a practice-based approach. Specifically, I identified a research article as an empirical contribution, when it included 1) a clearly stated methods section, regardless of whether it is a specific chapter or a part of another chapter with 2) a description or at least explicit notion of applied research methods and 3) an explicit case study or sample description. Theoretical essays as well as contributions presenting research synthesis were excluded from further analysis. To identify the kind of publication I screened the abstract and in the next iterative step the full text body for the references to empirical work. No articles, which stated in their title, abstract or methods section that they present a form of literature overview, were included. Further, domains such as health and medicine, tourism, and geography were excluded. After applying all inclusion and exclusion criteria in several iterative steps, the final sample for qualitative content analysis counted N=51.

Identifying empirical contributions for the sample was challenging in several ways. Most pressing was often very brief, vague, or lacking description of research designs applied in articles, which other authors conducting similar literature studies also notice (Kennedy et al., 2020, p. 44). For example, some authors provide an analysis for distinguished empirical cases either derived from previous empirical research, based on market analysis or omit explicit methodological details in their articles (e.g. Chan & Humphreys, 2018; Hartong & Piattoeva, 2021). In the core of such analysis lie analytical approximations to the role of datafication in a specific case, while empirical examples illustrate the contingencies of datafication therein. Although excluding such articles form my sample was required to focus on specific methods used to approach datafication, it illustrates one of the limitations to my sampling strategy that, in some cases, may not account for work framing methodology as inseparable from theory.

While theoretical and methodological articles or reviews represent a valuable contribution to my understanding of datafication scholarship in this study, they do not suffice the criteria of the 'method-in-practice'. All articles, which did not have a distinctive methods section, were also excluded even if they used empirical evidence to support their theoretical argument. All articles, which explicitly claimed to provide a new or adjusted framework, model, methodology or theory without referencing any empirical illustrations were excluded. There is a further limitation to these criteria, as they may as well exclude creative or artistic publications, which do not follow an established structure with clearly separated text sections. To compensate these limitations as much as possible and following the zooming in and zooming out approach (Nicolini, 2009b), I followed a strategy for the data analysis, which allows to focus on the broader context, and in the context of my thesis, I complemented the literature analysis with expert interviews with the authors of chosen articles. Finally, while some of the sampled articles were successfully retrieved in a search query, the subsequent analysis of full texts has shown that 'datafication' was only mentioned once or twice and not explicated: these publications were not considered for further analysis.

The first step of my literature analysis was zooming in on each single text at a time in a formal qualitative analysis (see Schreier, 2014). This part of analysis focused on answering the first research question on what methods are used in datafication research and how. A formal qualitative analysis is a part of the family of qualitative analysis methods (Braun & Clarke, 2006, 2019; Bryman & Burgess, 1994; Gläser & Laudel, 2010; Hsieh & Shannon, 2005; King & Brooks, 2016; Schreier, 2014). It sheds light on the formal criteria of the texts, arguments, or other types of data, rather than on the content itself. Thus, I consider it as an initial step of a qualitative content analysis, providing a formal description of the dataset and an orientation for possible groupings of sampled research articles. Although formal qualitative analysis originates from quantitative coding, I applied it as a technique for familiarization with the dataset and focused hereby on the deductive analytical categories. Among deductive categories were philosophical assumptions, research procedures and techniques (e.g. sampling), and researchers' "sociopolitical commitments" (Onwuegbuzie & Frels, 2016, p. 51). Identifying theories, epistemologies, and ontologies provided a baseline for situating the sampled literatures in particular strains of scholarship and contextualise the methods and the purposes of the study. Details of inquiry such as research scope, goals, questions, sampling strategies, and particular methods or techniques address the practicalities of the reported studies, while the issues of audience, outreach and funding give the readers hints on the authors' proclaimed values ((Greene, 2006) as referenced by (Onwuegbuzie & Frels, 2016, pp. 51–53)).

Complementary to the formal qualitative analysis, I conducted an exploratory quantitative and bibliometric analysis of the sample using research software VOSviewer (van Eck & Waltman, 2007) and R package bibliometrix (Aria & Cuccurullo, 2017). I complemented this quantitative exploration with further formal qualitative analysis (Schreier, 2014) of the sampled articles, while insights from the quantitative methods served as entry points for a deeper examination of the sampled publications. While an explorative analysis allowed to identify questions that authors of the sampled literature answer in different ways, a complementary qualitative content analysis allows to

examine the arguments made by the authors in more detail. Due to the sample size, the quantitative exploration was used for visualising the sample and as an entry point for a more detailed qualitative analysis, as I discuss in more detail in chapter 5. So, I conducted quantitative analyses, detailing 1) countries where the authors of sampled literatures are located according to their institutional affiliations and visualising 2) co-authorships in the sampled literatures in order to illustrate whether and how the sampled articles are connected to each other; 3) co-citations, in order to establish whether the sampled articles build on similar body of work on datafication; and 4) authors' keyword co-occurrence analysis in order to showcase what central topics datafication scholars themselves identify as important for their empirical research. Altogether, the visualisations produced through my quantitative analysis served for mapping out the (sub-)fields within the sample and creating an overview over topics relevant for datafication scholarship. I discuss the particular techniques and visualisations along with their results in the next chapter 5 of my thesis.

The combination of formal qualitative and quantitative analysis allowed to develop categories for mapping out the growing field of datafication scholarship and (critical) data studies. These categories encompass (1) terms related to datafication, (2) academic disciplines and fields in which datafication research is conducted, (3) societal domains in which datafication processes are taking place, (4) methods and theories used to understand datafication processes, (5) concerns addressing the implications of datafication processes, and finally, (6) understandings of what enables datafication processes. Chapter 5 of my thesis presents the results of the literature analysis and is structured according to these categories.[8]

4.3 Interviews with datafication scholars

To complement the literature analysis with the accounts of research practices—the methods assemblages in practice—and to conclude my synthesis, I conducted expert interviews with the authors of sampled articles. The aim was to inquire in more detail about the authors' methodological practices, choices, and their role in understanding datafication empirically. The sample for the interview was derived from the list of authors of analysed literatures. I contacted the first and or corresponding authors via email. In total, N=46 of authors were contacted with an interview request. The requests were sent in fall 2020, marked by the measures against the spread of the Covid-19 virus across the world ("lockdowns" resulting in distance modes of working, teaching, and learning). With universities and other institutions such as schools and kindergartens closed across the world, academic communities, alongside with other professional groups, faced challenges of online or hybrid teaching, home office combined with parenting or other caring responsibilities, other insecurities, and even losses of the dear ones to the Covid-19. These circumstances were acknowledged in the interview requests and became a reason to request participation in form of virtual expert interview or a written email response. The requests also included notes on research ethics and provided details on the ways of recording the interview. The initial interview request also contained the discussion topics. Due to the challenges of the pandemic everyone faced, both the email communication and the scheduling of the interview appointments were complicated and required effort and patience from all sides. The interest, support, and care of the interviewed experts towards my research project gave me additional encouragement for my doctoral journey.

Building the interview sample on the list of authors of the analysed publications, interviews were informed through their expertise and experiences from their empirical studies, professional, and institutional knowledges (see Blöbaum et al., 2016; Bogner & Menz, 2009; Gläser & Laudel, 2010). I understand the scholars' expertise as situated in their practices and not detached from these (see Dreyfus & Dreyfus, 2005), following a practice-based methodological approach. Out of overall 46 contacted authors, 29 agreed to participate. Initially, more interviews were scheduled but could

[8] For a systematic quantitative review of datafication scholarship including both empirical and conceptual work across different research domains, see Flensburg and Lomborg (2021).

not be conducted in the face of pandemic-related challenges. Additionally, three experts recommended by interviewees multiple times were recruited through snowball sampling to include more recent research and experts whose work informs datafication scholarship as represented through my sample, but who do not necessarily use the term 'datafication' themselves.

Due to the exceptional restriction measures during the Covid-19 pandemic and the geographical distance to the interviewees, all interviews were conducted virtually with video conferencing software preferred by each of the experts. In total, 30 interviews were conducted from October 2020 to February 2021. Two more experts provided written responses to the interview questions instead of a virtual interview. The interviews lasted between 30 and 60 minutes, with most interviews taking around 45 minutes. Most interviews were conducted in English, others—sometimes expectedly and sometimes not—in German. Informed consent was obtained. All interviews were audio-recorded with an open-source software OBS Studio[9] and transcribed. Some interviews were followed by informal talks with the experts. Before, during, and after the interviews and informal talks with the experts, I took notes inspired by the memos writing common for grounded theory (Strauss & Corbin, 1996). The interview transcripts and notes were anonymised and are used in the following in an aggregated form, individual quotes from the interviews used in my thesis were chosen so that they allow to preserve the experts' anonymity.

For the interview questions, I focused either on the analysed articles (co-)authored by the interviewees or on their most recent work on datafication, depending on the interviewees' preferences. The literature sample covers research articles from 2015 to early 2020 (online first publications appeared online late 2019), meaning that some of the empirical work done by the authors took place in the beginning of the 2010s. It is to assume, then, both the authors' careers, their personal understandings, and academic discourses on datafication developed further. Research projects and interests of the interviewed scholars may have significantly developed since. The interviews, therefore, take off with the discussions about the analysed publications, but also touch upon experts' other projects concerned with datafication processes, both previous, current to the date of the interview, and future ones. With that, interviews about past research topics also reflect the experts' current situatedness in the academic fields and in their careers. For the results presented in this chapter, it means that the results of the interview analysis not only deepen the understanding of current datafication scholarship developed through literature analysis in chapter 5, but also broaden this in regard to the scope of questions, empirical examples, and conceptualisations of datafication processes.

Elicitation interviewing techniques were used to understand the experts' personal, emotional experiences of their research practices. I continuously adapted interview questions according to the feedback and my 'feel' after each expert interview. Despite such preparations, the question whether I was actually able to capture methodological decision-making and research practices of the interview partners was accompanying me throughout the data collection. Going back and forth between conducting the interviews and reading about practice-based approaches allowed me to develop in my analysis a set of categories that reflected the manifold applications of methods assemblages in practice. In contrast to the synchronous interviews via video conferencing software, the written responses were given to the main interview questions and further elicitation questions (e.g. about surprising, annoying, or irritating aspects of datafication research) were not included. With that, written responses were shorter than interview transcripts and did not include as much personal, emotional experiences of the datafication experts as did the interviews. The written responses, nevertheless, provided me with important insights into the decision-making and research practices of both experts. Subsequently, the two given responses were anonymised and analysed together with the interview transcripts.

[9] https://obsproject.com/

One of the requirements for the expert interviews encompassed by the informed consent was preserving the anonymity of the scholars who shared their personal stories and details about their research projects. To live up to the obligation and promise of anonymity, some details about experts' projects that could be recognisable for an attentive reader keeping track of relevant literatures about datafication are anonymised in the following chapters. For anonymisation, I use categories of the next level of abstraction in relation to the word being anonymised: for example, in the hypothetical example of research discussing datafication in Germany, I would address the study as European or simply use the word 'country'. I also use words such as 'community' or 'practitioners' instead of identifiable definitions of professional or other groups addressed in the research projects of the interviewed experts. The omitted details of research projects that could make an expert identifiable served as a background knowledge.

Finally, I conducted qualitative content analysis (Saldaña, 2016) inspired by Law's (2004) definition of the methods assemblage as practices ordering different elements of the assemblage as "(a) present, (b) whatever is […] manifest in its absence; and (c) […] absent but […] other because, […] it is not […] manifest" (p.84). For that, I used qualitative data analysis software MAXQDA[10]. Following that practice-based methodological approach, I conducted open coding of the interviews, to develop categories that described both various research procedures and techniques, and people, objects, artifacts, things, and concepts being gathered together through the methods assemblages as practices. While identifying and categorising what elements of a methods assemblage are 'present' according to Law's definition, in the further iterative coding phases, I attended specifically to the differences between the developed analytical categories in order to account for what is 'manifest absent' or 'othered' —elements that each of the categories excluded. While 'methods assemblages' as an analytical category was derived from my research interest and questions, the codes that filled the category were developed inductively.

After several iterative coding phases, the *core category of 'methods assemblage'* consisting of three different methods assemblages was established empirically. The core category accounts for the coherence across each of the three methods assemblages in regard to the codes it includes on the kinds of the assemblage's elements. At the same time, the three methods assemblages differ among each other sufficiently. The three methods assemblages, discussed in detail in chapter 6 of my thesis, are (1) *the methods assemblage for analysing encounters with data representations*, (2) *the methods assemblage for tracing the dynamics of data infrastructures*, and (3) *the methods assemblage for reconstructing datafied regimes*. The methods assemblages, as chapters 5 and 6 show, build on the heuristic developed to map out datafication research in my literature synthesis, provide additional context to the literature analysis, and situate its findings in the lived experiences of interviewed datafication experts. In addition to the core category of methods assemblages, supporting categories were developed. Some of these, such as categories concerning *researchers' subjectivities* (e.g. career stage, personal standpoints towards methodological issues or datafication processes), *philosophical assumptions*, their *socio-political commitments* (e.g. participation in a certain scholarly community), *research procedures* (including individual techniques and methods of data collection and analysis) resemble categories used in the formal qualitative analysis of the sampled literatures. The categories, however, comprise inductively developed codes. Finally, another crucial supporting category concerns *concepts about datafication processes* applied alongside with the methods assemblages. Together with the core category of methods assemblages, this category allows answering the core research questions of my research project, as discussed in chapter 7. All supporting categories allow defining the three methods assemblages, their commonalities, differences, and allow building methodological sensitivity towards how methods assemblages drawing different elements together produce varying understandings of datafication processes.

[10] https://www.maxqda.com/

A difficulty and a limitation of the qualitative analysis presented here lie in interpreting projects and practices reported in the expert interviews based on my own positioning as a researcher and as a person. To avoid "the god trick" (Haraway, 1988, p. 581) in my interpretations, in the next sections, I elaborate in more detail on how my understandings of methods assemblages and datafication processes developed during my doctoral journey and how being an early career researcher in the same field I analysed in my doctoral project allowed a unique position for conducting research in practice. This and the previous chapters of my thesis serve to situate my own standpoint towards datafication processes and research methodologies, that is grounded in critical data studies, theoretical approaches to methods performativity as developed by Mol (2002), Law (2004), and Barad (2007), and practice-based methodology (Nicolini, 2009b; Schatzki, 2002). Additionally, arguments developed in sociology of science and feminist research traditions support the structure of my thesis.

4.4 Methods-sensitive inquiry

Research methods and practices of ordering these in methods assemblages are in the core of my empirical analysis and the reflection on the emerging field of (critical) data studies and datafication scholarship. The focus on methods assemblages, however, does not define this inquiry as methods-centric. Rather, I aim to describe how methods assemblages—including not only the method as a technique of data collection and analysis, but also the underlying ethico-onto-epistemological approaches, theories, and the research field-making by the scholars and other actors—produce sensitivities to the manifold of datafication processes and their heterogenous elements.

As Donna Haraway (2016) reminds us, "[i]t matters what stories make worlds, what worlds make stories." (p. 12) 'Staying with the trouble' of methodological performativity and paraphrasing ANT scholars' argument (e.g. Law & Singleton, 2013; Mol, 2010), I understand methods assemblages as *producing sensitivities to the kinds of worlds that partake in the production of knowledges about datafication processes*. Methods assemblages, then, can be understood as "a set of empirical interferences in the world, a worldly practice, or a lively craft" (Law & Singleton, 2013, p. 485) by performing which different kinds of heterogenous (human and non-human) actors are brought together to produce new knowledges and (datafied) empirical phenomena.

Various scholars have addressed methods' performativity not as a 'troublesome' challenge that interferes with production of knowledge about the world, but rather as a starting point and a basis for developing methodological sensitivities (Law, 2021; Law & Lin, 2020; Thompson, 2020). Addressing methods assemblages and their performativity, then, becomes not a question of how to deal with performativity, but a question of what we can learn from performativity of methods and how these knowledges allow us to understand datafication processes better and, also, from the perspectives of the marginalised, the oppressed, the *othered*?

Attending to datafication scholarship through the lens of methods assemblages, methods' performativity, and grounded in theories of practice, I developed iterative research design that allows mapping out different kinds of methods assemblages including techniques of data collection and analysis, the scholars' subjectivities, positions in their fields of study and academic communities, and the knowledges sought about datafication processes. My research design makes an emphasis on a multiplicity of methodological accounts used to study datafication processes by omitting the qual-quant divide. Rather, it focuses on how the concept of datafication can be explored and understood differently:

> "Translated into practice theoretical terms, this might mean that research participants ' conduct with different data production methods can be seen as social practitioners performing differently in and across practices as contexts. None of these performances a priori provides a more valid picture of everyday practices; the choices of methods should depend on the focus in the empirical research" (Halkier, 2017, p. 199).

In the beginning of this chapter, I quote Law and Urry (2004), who, following Haraway's argument that there is no innocence in the world, propose to think about two questions: "what realities do the current methods of social science help to enact or erode? And what realities might they help to bring into being or strengthen?" (p.396). With this chapter I attempted to outline my answers to these questions and describe a research design that aims to bring forth the multiplicities of theoretical, methodological, disciplinary, and empirical perspectives on datafication processes that are currently applied in social sciences. My thesis, therefore, not only follows the feminist, practice-based approach foregrounding methodological sensitivity to multiplicities, situatedness, otherness, and relationality. I address situations and develop categories concerning the ways in which researchers become aware[11] of the multiplicities, fragilities, materialities, and partialities (Law & Lin, 2020) of datafication processes and make methodological decisions on what to include how in their research practices. By elaborating and reflecting in the following chapters on these decision-making processes and the ways in that these enact different kinds of methods assemblages, I also aim to develop a heuristic outlining methodological sensitivities relevant for reflecting on the studies of datafication processes. In the concluding section of this chapter, I discuss my own standpoints and the social lives of my research methods. I also reflect on the impact of the Covid-19 pandemic on my research project and what role various datafication processes played in it.

4.5 The "social life" of my data inquiry during Covid-19 pandemic

The practice-based research requires from the researcher to become a practitioner themselves and engage with the studied practices ethnographically:

> "A well- known method for getting to know the rules and teleo-affective structures of the practices is to enter into the practice as a novice. Recognized new-comers to the practice are often granted the right to experiment, to ask basic or stupid questions, to find out about the expected doings and sayings, to learn how activities and projects are being carried out, and what kind of power relations are at stake" (Spaargaren et al., 2016, p. 18).

In my doctoral project aiming at understanding the performative role methods assemblages play in the datafication scholarship, I engaged with the practices I study as an early career researcher and a doctoral candidate working on a project situated within critical data studies and media studies. In this section, I describe my practical engagement and reflect on the social lives of my own research methods.

The experiences that are reflected in this section are the result of my positions as a doctoral candidate at the University of Bremen and a research associate affiliated with the research project DATAFIED[12] funded by the German Federal Ministry of Education and Research (BMBF; project funding number 01JD1803A) in which datafication processes are studied with particular focus on K-12 education in Germany. Being and working as a novice in academia, I, indeed, began getting to know 'the rules and teleo-affective structures' of research practice first-hand. I learned practices and challenges of empirical research, rules of grant applications, luxuries and challenges of interdisciplinary work, and everyday business of universities. Particularly as a member of consortium projects, together with my colleagues I started learning the care work of translating and developing common understandings of concepts across disciplines. Peer doctoral networks at the University of Bremen, within the DATAFIED research project, and those emerged from the participation in the virtual ECREA PhD summer school 2020 supported me by answering 'stupid

[11] In her book "If…then: algorithmic power and politics" Taina Bucher (2018) sets out to explore Facebook's algorithms and media by outlining situations in that people (for the first time) become aware of algorithms. Inspired by that analytical and rhetorical approach, I attempt to identify situations in academic scholarship in that researchers become aware of various aspects of datafication processes they study and make informed decisions about how to attend to these or not.

[12] https://datafied.de/

questions' and shedding light on the 'power relations at stake'. Working with other datafication scholars as a research associate, I learned hands-on 'how activities and projects are being carried out'. I also engaged in empirical datafication research myself, conducting empirical work with practitioners who enable and are impacted by datafication processes in German educational domain. Getting to know different disciplinary academic and practical perspectives on datafication processes, particularly in German school education, facilitated my understanding of datafication as socio-technical processes, in which digital data, infrastructures, and software are in recursive relation with the societal phenomena (Kitchin, 2014c; Jarke & Breiter, 2019). In the course of my PhD journey, this understanding also merged in developing methods (Jarke & Zakharova, forthcoming; Zakharova et al., 2022) and concepts (Zakharova & Jarke, 2022) for studies of datafication processes.

The beginning of the Covid-19 pandemic coincided with the planning phase of my empirical data collection. A move to 'everything' virtual required adjustments. Various resources (reading lists, blog posts, exchange with other colleagues) helped me to navigate the planning and conducting empirical research for this thesis. Particularly, reading lists such as crowd-sourced GoogleDoc moderated by Deborah Lupton[13] and the LSE digital ethnography reading list moderated by Zoë Glatt[14] were helpful for navigating virtual social research and the implications of the Covid-19 pandemic for social sciences (see Garcia & Barclay, n.d.; L. Taylor et al., 2020). As my interview partners sit in different universities across the globe, I was initially planning with virtual interviews, though with the pandemic it seemed inevitable and different. Even the interviews with scholars located closely to Bremen had to be planned as virtual meetings. When I conducted interviews several months into the pandemic, however, the experts and I were already used to virtual back-to-back meetings and collaboration; therefore, virtual interviews could be conducted without any major drawbacks. Nevertheless, conducting interviews virtually during a pandemic sometimes was challenging for everyone involved. The unstable internet connections on each side, time lags during the video conferences, time-consuming preparation of all relevant software and hardware by the interviewer before each interview, interruptions due to other, e.g. caring responsibilities of the interview partners all from time to time interfered with my schedule and plans for data collection. Scheduling appointments across time zones was unavoidable due to the locations of the interview partners, however during the pandemic it became more complicated. Some interviews had to be permanently cancelled, while I conducted some of the other interviews at 6am or 12pm local time from my living room in Bremen. Particularly due to all the additional challenges of the Covid-19 pandemic, the engagement and interest of the experts that I experienced played a significant role in moving forward my inquiry about and reflection on the emerging field of (critical) data studies.

Besides the interviews and in line with my practice-based methodological approach, I engaged in other activities relevant for datafication scholarship such as participating in academic conferences, writing, publishing, and other forms of knowledge exchange such as workshops and research colloquia. As the significant part of my doctoral project took place during the Covid-19 pandemic, several conferences have been cancelled, postponed, and those that took place were organised as virtual events. With that, as (nearly) everything during the pandemic, I could engage with my topic ethnographically mostly virtually. At the same time, conferences as virtual events allowed me more flexibility and possibilities to participate in more events than it would have been possible physically. Since 2019, I participated in seven international academic conferences (only one of which physically!) which at least some of the interviewed experts or the authors of analysed articles also visited. I attended their talks and myself presented results of the work conducted with

[13] https://docs.google.com/document/d/1clGjGABB2h2qbduTgfqribHmog9B6P0NvMgVuiHZCl8/edit
[14] https://zoeglatt.com/wp-content/uploads/2021/05/LSE-Digital-Ethnography-Collective-Reading-List-SHARED-DOC-May-2021.pdf

my colleagues. I engaged in discussions (as far as it was possible online) and read new publications published by the experts relevant to my analysis. I also engaged in academic writing beyond this dissertation and took the first steps in learning the practices and policies of academic writing and publishing. Finally, within the research consortia and working groups at the University of Bremen, I had opportunities to engage in further discussions and networks with other datafication scholars, also on social media. All this, alongside with the empirical and conceptual work conducted for this dissertation, shaped my understanding of both methods assemblages and datafication, thus also shaping the empirical findings presented in the following chapters. Particularly through writing within the DATAFIED research project, I also learned more closely some of the arguments from feminist lines of academic thought and adopted these for my thesis as well.

Being an early career researcher interested in datafication processes myself, I could not assume a position of an invisible observer. Rather, engaging in discussions with the experts during the interviews, engaging with their (and others') works on datafication processes through reading, writing, and talking created my partial connections within the webs and networks of datafication scholarship. Connections created through the interviews were literal: reaching out to different datafication experts, I could enter and peek into academic networks and learn their workings and rules. These connections were also personal—glancing into the lives of colleagues across the world through a little square on the screen and letting them (or not, as I mostly used the virtual backgrounds in video conference calls) to glance into mine (even though my cats still found a way into some of the interview transcripts). These connections were emotional: imagine talking in person for half an hour with a senior scholar whose work you admire or with a peer who has lived through similar situations you are facing. The connections developed in the interviews were also conceptual: they manifested in qualitative codes and categories interrelated through the concept of methods assemblages.

Finally, conducting expert interviews with other (often senior) scholars for my dissertation, I also learned from my interviewees the practices, challenges, and solutions to some practical, technological, ethical, and conceptual problems datafication scholars face in their work. Here, particularly in the interviews with senior colleagues, the interview, at times, resembled mentoring situations. For example, I4[15] noticed the research ethics documentation they received before the interview: "It is very good research work by the way. I wish all students were that well briefed about, you know, research ethics." Another expert commented on their understanding of datafication as research topic in my project:

> "Well, I guess, like, one of the challenges you might have is, what is datafication, what is not datafication in terms of this as a methodological approach. I guess that might be part of your research […]. So, I guess, at one point or another, there might have to be distinctions made about what kind of datafication we're talking about when we're talking about uses. But yeah, this is the comment." (I16, Pos. 29).

As the quotes suggest, my position of junior scholar and researcher was helpful for eliciting some well-known, unwritten rules of academic scholarship in general and datafication research in particular. Having received reflexive feedback after the interviews from some of the experts, I also hope that my questions and the interviews themselves also provided the experts with space for creating their own connections: remembering and rearticulating, rethinking, and reflecting their past, current, and future projects in the short time span of an expert interview. While in chapters 1-4 of my thesis I provided a theoretical and methodological background of my inquiry, in the chapters 5-8 I elaborate in more detail on my empirical findings and their implications for understanding the manifold of concepts about datafication and the field of data studies.

[15] Here and in the following, I will use the anonymised numeric codes for the interviews instead of the experts' names to retain anonymity of their accounts.

5 Mapping empirical data studies

For Latour,

> "[no] matter what we do, when we try to reconnect scientific objects with their aura, their crown, their web of associations, when we accompany them back to their gathering, we always appear to weaken them, not to strengthen their claim to reality. I know, I know, we are acting with the best intentions in the world, we want to add reality to scientific objects, but, inevitably, through a sort of tragic bias, we seem always to be subtracting some bit from it" (Latour, 2004, p. 231).

This chapter presents the results of a literature synthesis of original research articles from my sample of current datafication scholarship in social sciences. Through this synthesis, I explore how datafication scholars re-situate datafication processes they study in the empirical stie of practice and accompany 'scientific objects' 'back to their gathering'. In this chapter, I draw connections to the elements of the methods assemblages as defined in chapter 3.3: researchers personal, epistemological, and research-political positionings (as far as these can be addressed by reading their published work); the researched human and non-human actors; the sites of practice in which research processes described in sampled articles were distributed; research process, practices, and techniques used for academic knowledge production.

In this chapter, after providing an overview of the sampled research publications in section 5.1, I elaborate on various theoretical underpinnings of the term 'datafication' used in the sampled articles. The terms, concepts, and definitions discussed in this chapter stem from the analysed academic publications and are discussed here as means for understanding researchers' positionings in these publications as accomplishments of their empirical research practices. Specifically analysing how the authors of the sampled publications conceptualise datafication processes provides insights into the underpinning epistemological and ontological perspectives on the datafied society which the authors engage with in their work. The section 5.2 of this chapter draws on the chapter 2 of my thesis in regard to the different concepts and terms relevant for defining 'datafication processes' theoretically and empirically. The aim of the section 5.2, however, is to introduce theoretical underpinnings of the use of term 'datafication' in the sampled research articles. Moving further in my analysis, I show in the following sections of this chapter how the authors revisit and re-situate these theoretical conceptualisations through their own empirical research.

In section 5.3, I attend to the questions of what societal domains and empirical sites of practice are being studied in regard to datafication processes and what constitutes these datafication processes according to the sampled empirical articles. As my analysis is concerned with a sample of academic publications from various domains of social sciences (media studies, education research, information and communication technologies for development – ICT4D, among others), a detailed description of the variety of empirical sites reported in these publications would be repetitive to the analysed publications themselves and would hardly do justice to the work done by the authors. Therefore, when addressing the sites of practice discussed in the sampled articles, I focus on what the authors discuss as datafication processes as the common ground shared across the sample. By fleshing out the core elements of datafication processes discussed by the sampled articles' authors, I analyse what kinds of human and non-human actors are drawn together in the methods assemblages enacted by the authors.

Finally, I conclude with an analysis of matters of concern raised by the authors of the sampled publications in regard to the datafication processes they observed and methods used to study these. Both the matters of concern and the methods—particular techniques used by the authors for data collection and analysis—allow assumptions about the kinds of knowledges these authors sought about empirical datafication processes they researched. As I argued in chapter 3 of my thesis, these particular research techniques, although being a relevant part of the research processes, need to be analysed in relation to the other elements of the methods assemblages, if the performativity of these methods assemblages is studied. Section 5.4 of this chapter, therefore, presents my analysis of the empirical research processes, research designs, and goals, as reported in the sampled publications.

Based on these findings from the literature analysis, in this chapter, an initial overview over methods assemblages and multiple concepts of datafication developed and performed through the sampled empirical studies will be discussed. By analysing the literatures along the above listed aspects, I map the emerging field of data studies and develop a heuristic as a reflection tool sensitive and sensitising to the relation between scholars' positionings in their research and the multiplicity of empirical, re-situated conceptualisations of datafication processes (section 5.5). The analysis presented here is followed by an interview study with the authors of the sampled literatures, that completes my synthesis of empirical datafication scholarship in the chapter 6 of my thesis.

5.1 Mapping the literature

This chapter presents an analysis of 51 original research articles that report empirical findings on datafication processes from various domains of social research (see Appendix 1 for a full list of sampled publications). Overall, the analysed articles were published in 34 academic journals, ranging from communication, media, and cultural studies (14), education research (10), sociology (2), information systems (2), feminist studies journals (1), and outlets in that research across different disciplines such as social sciences, humanities, political science and information technologies research is published (5) (see Appendix 2 for the list of journals). Respectively, most of the sampled articles were published in outlets broadly covering various topics of media, communication, and cultural studies, covering topics such as activism and datafication processes (10), news, platforms and datafication processes (10), everyday practices and datafication processes (7). 18 from 51 further articles presented projects concerned with datafication of education (2 published not in specifically education-related outlets). The remaining articles covered issues such as governance in datafied societies (2, while further articles, e.g. from education research, also examined this topic), data justice and data infrastructures (4). Overall, the sample included articles published since 2015. As the sample included articles published online first up to the end of 2019, some of these have been published within respective journal issues in 2020 and received respective update of the publishing date in my sample (6, see figure 5-1). The most articles per year were published in 2019 (19). The yearly growing number of publications reflects the expanding use of the term 'datafication' in social sciences and related disciplines, illustrated in the previous chapter.

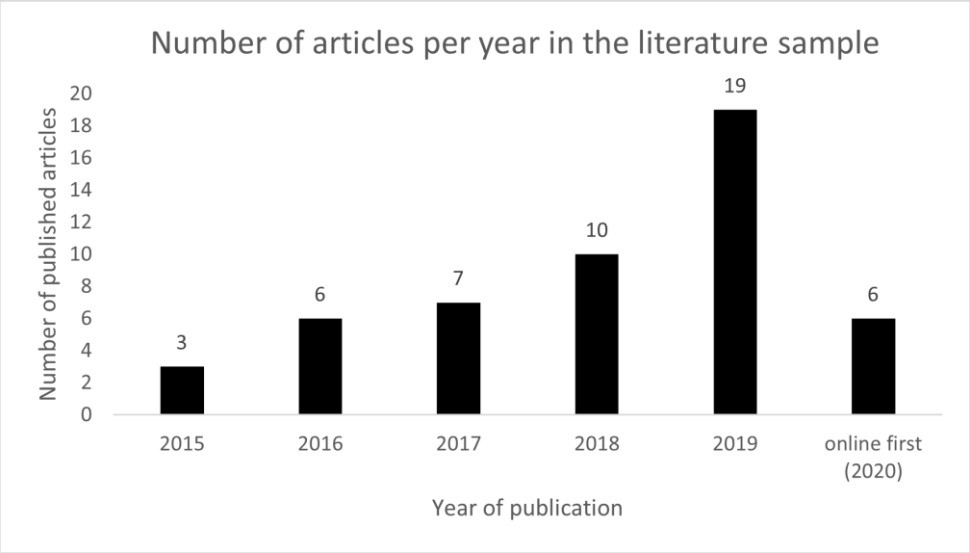

Figure 5-1 Number of articles published each year in the literature sample.

I conducted a formal qualitative analysis of the sampled articles to further familiarise myself with the sampled literature and ensure the heterogeneity of the collected sample. First, I attended to the geographic heterogeneity of sample. I examined how presented research on datafication covers several countries and geographical regions, according to the authors' institutional affiliations. Table 5-1 lists the countries of affiliated institutions (at the time of publishing). As it is typical for at least some social sciences research domains (e.g. Waisbord, 2019) and because only articles in English language were included in the sample, most authors are affiliated with universities in northern America and Europe. Sometimes research reported in the articles was conducted in regions other than scholars' residence regions. For example, Kelly and Noonan [24][16], Masiero and Das [32], and Taylor and Richter [47] report on empirical research in India, while Halkort [14] reports on her research in Palestinian refugee camps. Heeks and Shekhar [18] develop an applied data justice framework supported by evidence from local initiatives in Kenya, India, and Indonesia. Lee [26] analyses social credit score system in China, while Chen and Qui [8] focus on platformisation of Chinese urban transport. Candido [7] reports on datafication of education quality assessment in Brazil, while Piattoeva [36] presents insights into the datafication and surveillance in Russian educational domain.

Table 5-1 List of countries of the authors' affiliated institutions (at the time of publication)

Australia	Hungary	New Zealand
Canada	India	Norway
Denmark	Ireland	Switzerland
Finland	Italy	United Kingdom
Germany	Lebanon	United States
Hong Kong	Netherlands	

Further, within my formal literature analysis, I used bibliographic coupling analysis to examine whether the authors of sampled publications have shared theoretical backgrounds (based on references used in the sampled academic articles). Figures 5-2 were created with VOSviewer. The nodes illustrate authors, while each node's size is weighted according to the number of articles

[16] In order to distinguish between the references to the academic publications in my sample from other references used in my thesis, in the literature analysis (chapter 5), the sampled articles will be referred to according to their number in an alphabetic list presented in Appendix 1 of my thesis. Appendix 1 includes the full reference to each of the sampled articles.

in the sample published by each author. The edges between the nodes are weighted and illustrate the overlaps in references between each two authors. The colours were assigned to the nodes automatically and indicate clusters. An analysis of each coloured cluster reflects domain-specific and/or thematic communities of which each scholar is part.

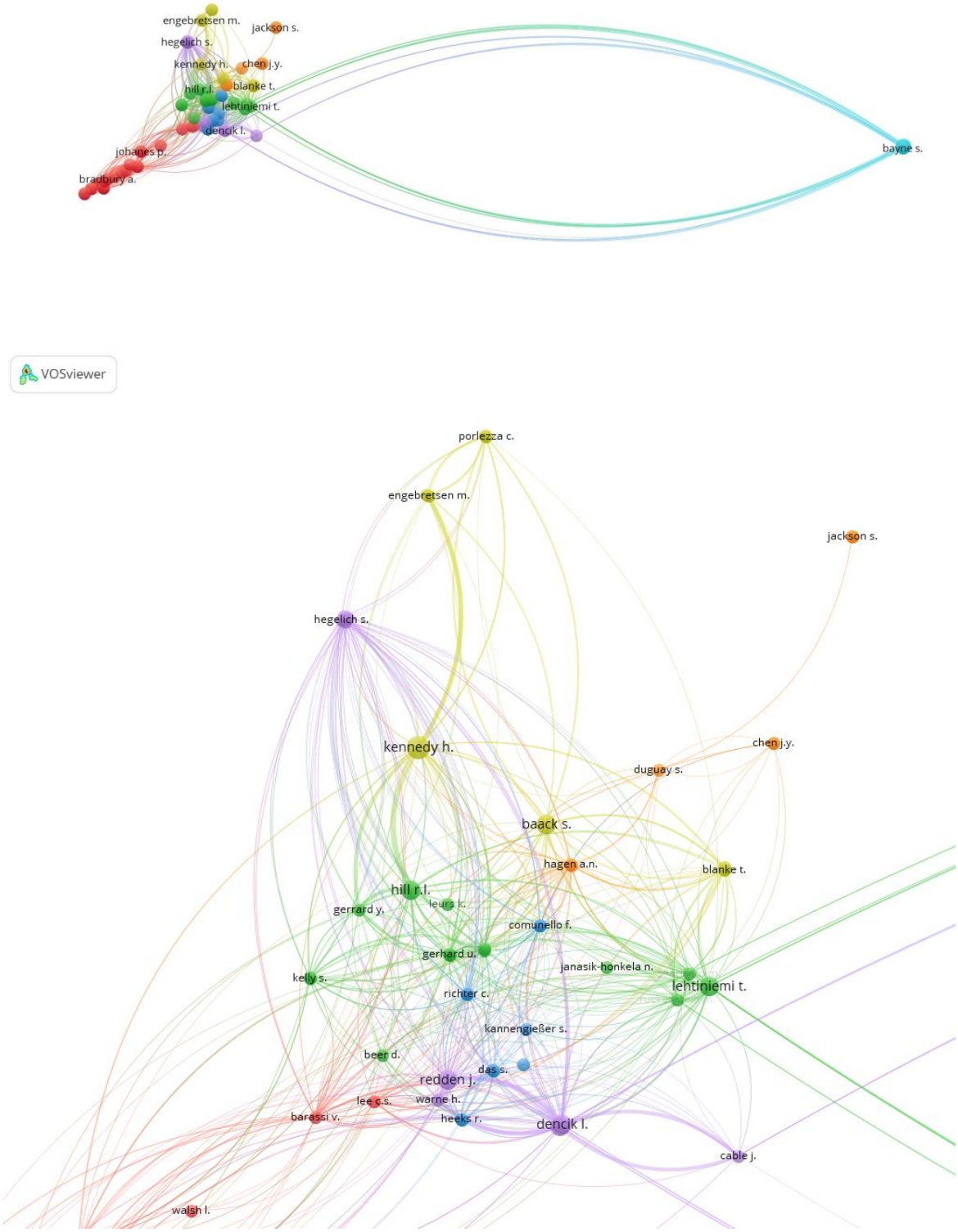

Figure 5-2 Visualisations of co-authorships in the sampled literatures. Created with VOSviewer. Visualisation of all co-authorship associations (above) and a snippet for better readability of the smaller clusters (below).

For example, figures 5-2 illustrate that there are several thematic and/or domain-specific clusters connected to each other (e.g. red cluster can be seen as education research, purple cluster as studies using the data justice framework, and light green—as studies of news media). These clusters also reflect the domains covered by the journals in which sampled research articles have been published. As figures 5-2 illustrate, the colour-coded clusters are not separated from each other, rather multiple connections between these clusters can be seen according to the bibliographic coupling analysis. Data studies, then, explored in regard to the empirical research on datafication as represented through the literature sample in my thesis, can be considered as cutting across and building on multiple research domains within social sciences. It can be assumed that data studies provide an analytical and empirical common ground for researchers in media and communication studies, education research, as well as governance and activism scholars. The close positions between the nodes indicate shared theoretical background of the authors across the domain-specific clusters within social sciences (van Eck & Waltman, 2007) and allow viewing datafication scholarship sampled here as an indicator of a rather coherent academic field. This kind of analysis, however, particularly taking into consideration the size of the analysed sample, does not allow assumptions on how exactly concepts and knowledges travel between the research domains and fields, where these concepts and knowledges originate and to what extent are they imported. To better understand how data studies are composed according to the analysed sample, further analysis is required.

So, in order to determine which academic publications constitute the shared theoretical background of the sampled publications, I analysed the sample according to co-citations. Examining co-citations within the sample allows to identify, which further scholars were cited by the authors of the 51 sampled publications. Figure 5-3, also created in VOSviewer, illustrates the authors' co-citation network. Each node represents a scholar whose work is referenced in the sample. The size of the node indicates the number of references in the sample. Proximity of nodes to each other indicates that the publications by these scholars are cited together. Co-citation analysis, however, does not distinguish between the references the authors build on and those they critique in the analysed publications. For that, more detailed analysis is required. The section 5.2 of this thesis, dedicated to situating the use of the term "datafication" in the sampled research articles, provides some of such analysis, specifically discussing conceptualisations of datafication used by the authors and whether these conceptualisations are embraced or critiqued. Figure 5-3 shows that academic contributions by Stephen Ball, Alice Bradbury, Nick Couldry, Kate Crawford, Rob Kitchin, Bob Lingard, Deborah Lupton, Jenny Ozga, Sam Sellar, Jose van Dijck, Ben Williamson (alphabetic order) were referenced in the sample 20 times or more each, works by Marc Andrejevic, Kenneth Cukier, Helen Kennedy, and Viktor Mayer-Schönberger were also cited nearly as often. The co-citation analysis also illustrates differences between the clusters that can be addressed as education research (on the right-hand side of the visualisation) and other clusters that can be broadly described as media studies, communication research, cultural studies. While these clusters reflect the sample, the co-citation analysis presented through figure 5-3 further expands the scope of research relevant to data studies beyond the sampled literature. Figure 5-3 draws connections to research excluded from the sample, such as e.g. human geography (represented in the figure through references to the works by Rob Kitchin) or research on health (e.g. represented here through references to the works by Deborah Lupton). The presence of these references, thus, indicates that a different literature sample, e.g. one including other research domains excluded from my analysis such as health and geography, would comprise these further clusters under the umbrella of data studies as well in addition to the clusters indicated in figure 5-3.

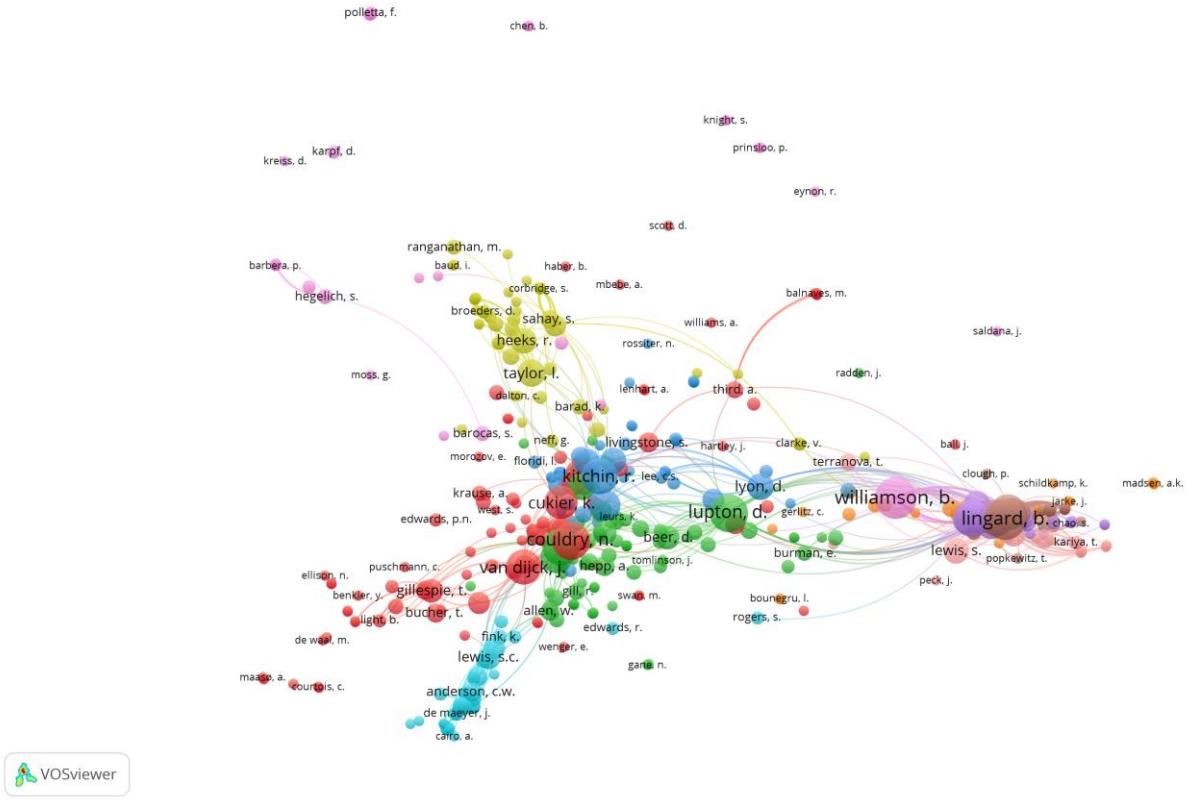

Figure 5-3 Visualisation of co-citations in the sampled literatures. Created with VOSviewer.

In this context, to understand better the role data studies play in associating these heterogenous research themes and domains, some of the most frequently shared references within my sample of publications are noteworthy. So, unsurprisingly, the seminal text by Mayer-Schönberger and Cukier (2013) was among the most cited ones (as my analysis in the section 5.2 will show, however, this work was also critiqued by several authors of the sampled articles). Besides this seminal text, publications by van Dijck (2014), Kitchin (2014c), and boyd and Crawford (2012) were the ones cited most frequently (>6 times) in the sample. These authors are also among the most frequently cited scholars in my sample. As further analysis of cited references indicates, among the works of these authors the mentioned seminal texts are cited nearly exclusively. For other authors identified in figure 5-3 as frequently referred to in the sample, a bigger number of publications with a smaller number of citations each could be identified. Drawing on these findings, the mentioned seminal texts can be considered as providing the conceptual foundation for bringing studies in different disciplinary and thematic domains together.

While the bibliographic coupling and co-citation analyses indicate some shared theoretical backgrounds on datafication, further analysis of the sampled articles also shows methodological coherency. So, most scholars whose publications were included in the sample apply qualitative techniques of data collection and analysis (40 articles out of 51) across various academic fields and domains covered in the sample. Ethnographically inspired observations and interviews (or focus groups) coupled with various kinds of text analysis (incl. policy research) are among the most commonly applied data collection and analysis techniques in the sampled articles. Other qualitative methods such as media diaries and self-monitoring [25, 45], story completion methods [26], and theory-inspired methods [37, 41] were also applied. In a few other articles, the use of various methods [4, 27, 48, 49], both quantitative and qualitative, was reported in one or multiple projects. In the latter case these projects were conducted over a longer period of time [6, 44, 50], while two other articles report about multiple projects in that primarily different qualitative techniques were

applied [24, 44]. In some further articles, participatory and action research-inflected studies were presented [28, 29, 39]. Only one article in the sample reports results from a solely quantitative, computational study [34]. Finally, digital methods such as digital ethnographies or walkthroughs in combination with other qualitative methods such as interviews and observations, were also used [3, 11, 21].

In order to explore the sampled literature further, additionally a keyword co-occurrence analysis of author keywords was conducted with VOSviewer (figure 5-4). The keyword "datafication" and its wordforms used as a search term for the identification of the sample was removed from the co-occurrence analysis. Overall, 204 different keywords were analysed. The keywords listed by the authors of the analysed articles and printed in the beginning of article were extracted automatically by VOSviewer (van Eck & Waltman, 2007) from the exported database search query (in VOSViewer, "author keywords" analysis was selected). Two articles did not have any keywords [22, 28]. In figure 5-4, author keywords are represented as nodes, the size of nodes symbolises the number of times each keyword is used in the sample. The connections between the nodes illustrate which keywords co-occur—are mentioned together. The thicker the edge between the nodes, the more often the keywords are mentioned together. The colours of the keywords were assigned by the VOSvieiwer based on the proximity of the keywords to each other. Nine smaller clusters on the sides of the graph illustrate the keywords used together in a small number of articles, with no further connections to others.

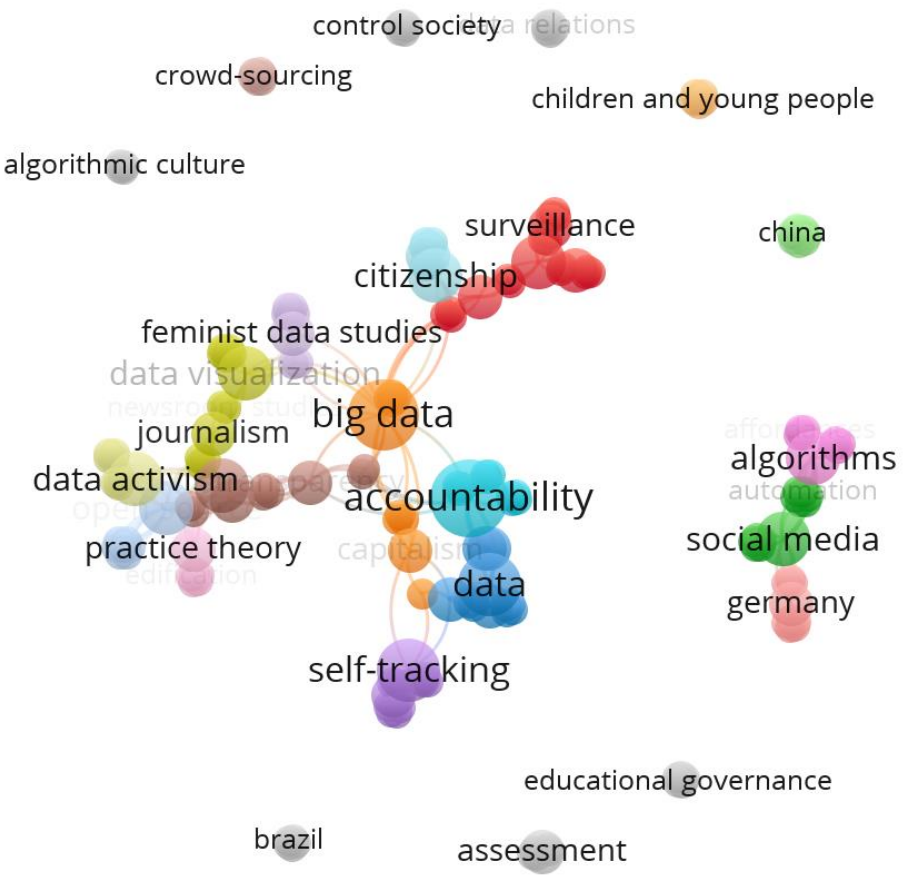

Figure 5-4 Keyword analysis of the literature sample. Visualisation generated with VOSviewer. The visualisation illustrates keywords used multiple times, some keywords used only once are invisible in this depiction.

Figure 5-4 shows that a number of similar keywords such as e.g. data activism and activism or visualisation and data visualisation are not connected to each other. In the data cleaning process, these keywords have not been merged on purpose to identify how the sampled literature can be clustered. The gaps between such similar keywords illustrated through the analysis allow for two following assumptions. First, there is a density of terms and concepts used to describe und conceptualise datafication processes across academic domains and fields. Repeating keywords indicate that scholars across different domains are concerned with similar datafication-related issues. Secondly, however, a number of similar but distinct keywords used by authors indicate that the empirical datafication research is yet not consolidated, despite some shared theoretical backgrounds. Rather, there are different clusters, that may illustrate various academic communities that apply varying epistemologies and methodologies to the studies on datafication. The remainder of this section interrogates that assumption.

A close look at the keywords used in the sampled research articles suggests that these allow for contextualising the literature on several levels. Overall, after exclusion of search terms, 204 remaining keywords were analysed. Of these, most were only used once. Among the most frequently used keywords were 'accountability' (6), 'big data' (5), 'self-tracking' (4), 'data' (4), 'algorithms' (3), 'citizenship' (3), 'data activism' (3), 'data justice' (3), 'data visualization' (3), 'open data' (3), 'open source' (3), 'social media' (3). These keywords indicate different societal domains impacted by datafication (e.g. 'data activism', 'social media'), various concerns arising due to datafication of these domains (e.g. 'accountability', 'citizenship'), and phenomena that are being described as datafication (e.g. 'algorithms'), as well as methods and techniques used to make sense of datafication in research and practice (e.g. 'data visualisations'). While the most frequently used keywords point to the possible thematic clusters among the analysed texts, a further analysis of these clusters is required.

First, keyword analysis identifies a number of *terms and concepts, related to datafication*, like 'big data', 'digital traces', 'metrics', 'assessment', 'quantification', 'automation', 'commodification', and 'digitalisation'. These keywords represent other academic discourses, closely related to the concept of 'datafication' according to the authors. Importantly, these keywords also represent different conceptual and epistemological approaches to datafication. Section 5.2 of this chapter attends in more detail to the manifold of terms and concepts related to datafication based on the content analysis of research articles and puts 'datafication' in the context of other academic discourses addressed in the sampled research articles.

Second, some keywords indicate the *societal domains (also including geographic regions) where datafication processes are being enacted and studied*, such as politics, 'data activism', 'journalism', and 'social media'. In contrast to academic disciplines and fields, these keywords indicate a manifold of answers to the question of what is being datafied in our societies. The above listed examples of such keywords imply the bandwidth of datafication processes that envelope both individual practices and experiences (e.g. 'emotions'), groups (e.g. 'youth', 'hacking culture'), public institutions (e.g. 'public administration', 'school actors'), and commercial organisations (e.g. 'facebook'). Since there is a multiplicity of definitions of datafication used across various academic domains and discourses, it is not surprising that for different scholars the concept of datafication is meant to describe a variety of processes and technologies. Accordingly, some keywords such as 'algorithms', 'anti-poverty programmes', 'civic technology', 'national testing' [of schoolchildren], 'data analytics industry', or 'quantified self' all illustrate the examples of *what scholars are talking about when talking about empirical datafication phenomena.* As the listed examples of keywords indicate, different conceptualisations of datafication describe, in some cases, a manifold of technologies (e.g. 'algorithms' or 'civic technology'). For other scholars, datafication stands for national policies and political initiatives largely based on data infrastructures and aiming at data-driven decision making (e.g. 'anti-poverty programmes' or 'national testing'). Further, datafication can address lived experiences of particular individuals or groups as indicated by e.g. 'quantified self' keyword, or

situate the role of data for creating economic value (e.g. 'data analytics industry'). Section 5.3 interrogates the questions of what is being datafied and what the term 'datafication' stands for in the sampled research articles.

Third, some keywords refer to matters and methods of *concern that datafication scholars attend to empirically*. Some keywords explicitly address research methodologies and methods applied to explore datafication processes, such as 'feminist data studies' or theoretical frameworks used to explain these processes such as e.g. 'practice theory' (see figure 5-4). The matters of concern indicate some core issues negotiated in democratic, capitalist societies and include 'accountability', 'inequality', 'citizenship', 'surveillance' (and 'dataveillance'), 'transparency', and 'agency'. These concerns relate to the terms used alongside with 'datafication' and recount various kinds of relations between datafication processes and societies. Finally, these concerns and the conceptualisations of empirical datafication processes they help to develop are closely related to the research methods and techniques datafication scholars apply in their empirical studies. Section 5.4 of this chapter, therefore, attends to the methods and matters of concern that are addressed in the analysed literature.

Although the quantitative exploration of my literature sample indicates a manifold of concepts and methodological approaches applied to study datafication processes, it also illustrates coherency, for example in regard to the central literatures introducing the term 'datafication', the methods applied, and the concerns scholars across different disciplines raise. Therefore, I argue here that my analysis of empirical publications reporting research on datafication provides a glimpse into the current process in which data studies defined more broadly are being established as a research field. In the next sections of this chapter 5, I elaborate in more detail how this emerging field is constituted conceptually, empirically, and methodologically.

5.2 Situating datafication conceptually

The keyword co-occurrence analysis in the previous section indicated multiplicity of terms used by scholars alongside or synonymously with datafication. Keywords such as 'big data' and 'digital traces', 'metrics', 'assessment', 'quantification', 'automation', and 'commodification' can be seen as terms and concepts related to datafication, for example enabling datafication processes or being enabled by these. This section of my thesis explores the listed keywords above and their relations to the 'datafication'. Even the term 'datafication', however is differently defined by various authors referring to various seminal literatures on datafication and digital data. Overall, across the sampled literatures, several seminal definitions of datafication are quoted such as those developed by Mayer-Schönberger and Cukier [referred to by 1, 10, 12, 19, 23, 25, 27, 30, 45], van Dijck and colleagues [referred to by 2, 5, 11, 12, 20, 26, 28, 30, 45], Couldry and colleagues [referred to by 3, 12, 14, 30], as well as Bradbury and Roberts-Holmes and Williamson and colleagues in the contributions on datafication of education [e.g. 6, 7, 31, 37, 46, 50]. As the listing of authors referring to different conceptualisations of datafication illustrates, several authors offered in their publications multiple seminal definitions of datafication. These definitions, in many of the analysed articles, were further adapted and situated based on the authors' empirical findings. These empirically grounded concepts of datafication are discussed further in this section 5.2. Besides definitions of the term 'datafication', authors of the sampled literatures also referred to seminal works from critical data studies and other research domains in order to define digital data, digital traces, data assemblages, and data infrastructures: in addition to the already mentioned above scholars, works by Marc Andrejevic, David Beer and colleagues, danah boyd and Kate Crawford, John Cheney-Lippold, Nick Couldry and Andreas Hepp (e.g. referring to digital traces), Lisa Gitelman and colleagues, Rob Kitchin and colleagues (referring to data assemblages), and others, were cited. Finally, some authors also addressed other terms such as platforms [e.g. 8, 25, 27, 35]. In the following, these seminal conceptualisations of datafication, the aspects of datafication processes these definitions highlight,

and their relations to other concepts and academic discourses are discussed. While some of these conceptualisations referred to by the authors of the sampled research articles have been mentioned in chapter 2 of my thesis, the goal of this chapter 5 is different. In chapter 2, I provided an overview over the current state of research, while chapter 5 makes an inquiry into various elements of the methods assemblages, as they are discussed in the analysed publications. This section 5.2 is specifically exploring conceptual, epistemological underpinnings of the term datafication and this term's relations to others, as far is these can be studied through the literature analysis.

5.2.1 Datafication and quantification

For many scholars, datafication is closely related to quantification. This relation is made particularly visible in the definition of datafication offered by Mayer-Schönberger and Cukier (2013), which is referenced in many of the analysed articles [see 1, 10, 12, 19, 23, 25, 27, 30, 32, 34, 35, 39, 45]. According to this definition, datafication can be understood as quantification of an increasing number of societal aspects and phenomena: "the ubiquitous quantification of social life (Mayer-Schönberger and Cukier, 2013: 78)" [cited by 1, p. 2]. Although widely quoted as an initial conceptualisation of datafication processes, the definition of datafication by Mayer-Schönberger and Cukier (2013) is also contested in multiple sampled articles[10, 25]. For example, this critique is articulated by Kennedy and Hill [25] through the connection of that definition to the

> "rhetoric about big data's promise – as seen, for example, in Wired editor Chris Anderson's (2008) widely cited assertion that 'with enough data, the numbers speak for themselves' – to produce a faith in numbers" (p. 832).

This critique illuminates how the enthusiasm around digital data and computational methods of analysis (both in research and practice) brackets out the context where these data are situated, produced, processed, and used. This conceptualisation of datafication through counting and rendering into data various elements of the social world (Mayer-Schönberger & Cukier, 2013), however, relates datafication to quantification, commercialisation, and value extraction, also spreading across societal domains alongside with datafication processes. In this regard, datafication and quantification are put in relation to each other as societal processes with broad implications for various societal domains.

Masiero and Das[32], in turn, showcase how historical origins of quantification, in particular for the Global South, are connected to the history of colonialism. They, therefore, suggest viewing quantification and datafication as a continuation of colonial practices of counting and surveying, previously conducted with the use of e.g. census data, and currently perpetuated as demographic and other characteristics of people are more easily put in relation to, interconnected with each other through computational means (p.930). In a similar context of Indian datafied water system, Taylor and Richter [47, p. 727] also interrogate the question of 'what counts' politically and ethically. Both Masiero and Das [32] and Taylor and Richter [47] underscore the double meaning of the question of 'who counts'. On the one hand, it addresses the actors who do the counting and make decisions of whether or not and how certain empirical phenomena should be conveyed through data infrastructures, and ultimately into quantifiable data. On the other hand, 'who counts' is an ethical and political question of who is included and excluded, who is rendered (in)visible through the counting. For example, Taylor and Richter [47, p. 727-728] show how different actors involved in the datafied water system in the city of Bangalore in India have diverging views on what counts as (equitable) water supply. These views are not only diverging, but essentially quite opposite, as the boards and companies responsible for water supply identify as 'leakage' what citizens perceive as their access to water. Winter [51, p. 57] addresses the question of who counts in the context of educational governance, related to categories assigned to people through data. In the context of education, the consequences of such categorisation are particularly illuminating. Interrogating the relation between datafication and counting methods and measures used in educational governance, Winter [51] points out how the categorised subjects and not the

frameworks of categorisation and measurement become "amenable to regulation and control" (p.57). Other educational scholars [e.g. 7] also underscore how through datafication, the measurement frameworks become hidden and invisible, while the subject and object positions meant to be represented through data are rendered visible. Pybus and colleagues [39] make a related methodological argument discussing how "empirical analysis is transformed into an automated algorithmic effect as subsequent predictive analytics become fact" (p. 3). The quote illustrates how the categorisations created by certain actors holding power over regimes inscribed in software and data infrastructures often remain unquestioned, hidden, and become facts as soon as predictions based on these categorisations are made. At the same time, Bradbury [6] notices that in the context of education some phenomena are more suitable for quantification than others "[l]iteracy and maths are more easily translated into numbers" (p. 16). Human behaviour and especially human relationships, emotions, or more abstract concepts such as beauty [35] or friendships [25] are more difficult to quantify. In this regard, quantification and datafication can be both understood as detaching numbers and data from the context in that these data are being produced. Instead of illustrating diverging needs, data and numbers provide accounts about social realities, such as water required, supplied, and leaked as in the contribution by Taylor and Richter [47]; these accounts, then, are put to use for governance or accounting. In this first broad cluster of conceptualisations of datafication processes according to the analysed research publications, datafication is elaborated in terms of its 'impact' and implications on individuals and society, further reifying the division between operations of technologies and data on the one hand, and the social actors finding themselves at the mercy of these technologies/data. At the same time, one of the central aspects of digital data as defined by Borgman (2015)—their ability to be put to use for various purposes—is highlighted. Central for conceptualising datafication becomes the question, what are the ends of empirical datafied phenomena.

In the analysed literatures, datafication is also put in a relation to other concepts and terms such as quantification or digital traces through the ways of data generation and affordances of digital data such as machine readability or "interconnectivity" (Günther et al., 2017, p. 201). These affordances, alongside with a greater numbers of data points, allow for phenomena that previously were rather difficult to assess to be analysed more easily. In relation to datafication, quantification, thus, addresses the use of computational, algorithmic quantitative methods both in research and at the empirical sites of practice to make sense of digital data that are being produced in various societal domains. As some authors notice [25, 44], quantification and the role of numbers in knowledge-production have been widely studied, in particular in sociology, over a long period of time. So, referring to Porter (1995), Kennedy and Hill [25, p. 831] recount his arguments about faith in numbers that are equally relevant for digital data, addressing challenges that laypersons are facing being exposed to these numbers and the necessity of understanding and interpretation of these. Furthermore, as Engebretsen and colleagues [12] argue, datafication understood through quantification also allow "an increased accessibility of public data and easier ways of mediating them through a growing array of visualization tools." (p. 1) At the same time, digital data can be produced within the material infrastructure, for example through sensors [e.g. 32]. For some of the authors of the analysed research articles, data, therefore, do not only acquire material properties of the goods such as water, distributed through the infrastructures, but as new agents also change and create new relations. At least in the context of infrastructure studies, datafication, thus, draws on power generated and available through the materiality of infrastructures and goods turned into data within these. Infrastructures and goods are one example of the phenomena that allow, enable quantification and datafication processes. Quantification, in that context, takes place in accordance with existing measurement standards [for example, meters of water pipes in the example by 47]. Through such conceptualisations, rather technological aspects of datafication are being put forward, while digital data, their properties, and applicability for practical use —for some of the

authors—and technologies or infrastructures—for others—become central objects of their empirical studies.

5.2.2 Datafication, technologies, and experiences with these

Conceptualisations of datafication processes developed by Couldry and colleagues (e.g. also Couldry & Hepp, 2017) are expanded on by some authors of the sampled literatures [see e.g. 1, 3, 12, 15, 30]. In this research, datafication is specifically understood through practices and technologies [e.g. 3] or as a part of 'deep mediatisation' (Hepp, 2016) [referred to in 13]. These authors are particularly interested in one form of digital data, conceptualised as digital traces for example by Couldry and Hepp [13], to whose work Hand and Gorea [15, p. 669] refer, proposing

> "if digital traces are derived from "disparate kinds of data that are generated by our practices in a digital media environment" (Couldry & Hepp, 2017, p. 162), […] it is imperative to understand how data are generated by individuals, and what the relations between everyday practices, devices, and data are."

This quote also illustrates the connection between digital traces, empirical practices and technologies that are central for the understanding of datafication discussed in this paragraph. For Hand and Gorea [15] digital traces can be exemplified as "e-mails, texts, tweets, and tags visible in social media but also obscured locational, transactional, and temporal metadata" [p. 666, see also 1, 13]. Baack [2] further points out how digital traces left by people "are often unconscious and not meaningful to them" (p.2) and detached from the lived experiences.

The definition of datafication by Williamson as "ways of seeing, understanding and engaging with the world through data" [7, p. 129] further illustrates how human behaviour and methods used to render these data accessible for further analysis are being datafied [31, 46]. Similarly, Pierlejewski [37] expands on the definition of datafication developed by Cheney-Lippold as "the transformation of part, if not most of our lives into computable data" (Cheney-Lippold, 2017, p. 9) [cited by 37, p. 2]. Quantified self may be viewed as another prominent example where behavioural data become quantified for different reasons and in order to achieve a variety of goals. Parisi and Comunello [35, p. 69] theoretically explore the multiplicity of motivations guiding individuals to render their behaviours assessable to the quantitative analysis by creating digital data about these behaviours. Gerhard and Hepp [13] elaborate on the recursivity of data in context of quantified self as they draw a circle of "discursively contextualized practices that are deeply entangled with the respective tracking repertoire", where data offer feedback to the practices, orientation for desired practices, and execute control required to reach desired data (p.691). The idea of digital traces (as developed by Couldry & Hepp, 2017 e.g. referred to by [13]) illustrates this relation of an individual's behaviour "and subsequent software-based analysis by means of apps and online platforms" [13, p. 684]. However, as Baack [1] points out in the context of data activism, affordances of digital data formats do not necessarily have to lead to negative consequences, but may as well allow for more agency, transparency, and "creating 'actually existing alternatives' to established forms of knowledge production and circulation" (p. 8). Nevertheless, critique of datafication processes is prevailing in the sample of analysed literatures. Similar to how Gerhard and Hepp [13] describe quantified self, Leurs [29] addresses data mining as partial, subjective, and "power-ridden" (p.133).

Finally, close to a definition of datafication through quantification of everything is another conceptualisation, e.g. offered in my literature sample by Heeks and Shekhar [18]. It attends to the "growing volume, velocity, variety and visibility of data, with greater use of new forms and streams of data in decision-making (Heeks, 2018)" [18, p. 993]. Connecting this conceptual understanding of datafication with possibilities of empowerment as well as risks of growing inequalities, the authors not only view datafication as a technological process, but also address its implications on various elements of a data assemblage [18, p. 995]. As discussed in chapter 2 of this thesis, some concepts of datafication focus primarily on technological processes enabling operations with greater

quantities of digital data in shorter periods of time. While, for example, the latter quote illustrates an application of such a conceptualisation, both the authors of this contribution and others view these technological possibilities as situated more broadly in the society. In the sampled literatures, even when used definitions of datafication foreground technological possibilities of datafied infrastructures such as increase in volume, velocity, variety, and veracity of data (Gitelman, 2013), either various societal implications of such technological processes are foregrounded, or the authors also offer explicit and implicit critique of such technological understandings [see e.g. 13, p. 688]. Within this cluster of conceptualisations, datafication processes are, on the one hand, centrally addressed through technological transformation underpinning these processes: material and data infrastructures, in which data are being generated either through sensor devices or through user behaviour, digital data moving across these infrastructures. On the other hand, these technologies and digital data are addressed in their interrelation with various actors, whose behaviours, practices, and experiences—not at least with the said technologies—are being rendered into data.

5.2.3 Datafication and economic issues

Other scholars turn to inherently social and critical conceptualisations of datafication developed by van Dijck and colleagues (e.g. van Dijck, 2014; van Dijck & Poell, 2013) that foreground the complex power relations, commercialisation, and commodification developing, alongside, through, and within empirical datafication processes [see 2, 5, 11, 12, 20, 26, 28, 30, 45]. For example, building on van Dijck and Poell (2013, pp. 9–10), Duguay [11] addresses datafication as "the rendering of social media activity into commodifiable data, presented as a raw representation of popular opinion in real time" (p. 22), thus relating datafication processes to commercialisation and for-profit value extraction from digital data. Similarly, Jackson and Kuehn [20], also quoting van Dijck (2014), address datafication as "the collection, storage, usage and/or sale of user data for the purposes of profit generation" (p. 422). Thus, besides quantification, another concept closely related to datafication in the analysed sample, is commercialisation. In contrast to previously discussed conceptualisations, commercialisation clearly addresses economy, its role in various aspects of social life, and the ways digital data and datafication processes are relevant for both. Previous paragraphs addressing datafication technologically (as greater volume, velocity, veracity) and through quantification (data as machine-readable) illustrated how digital data can be more easily, faster, more efficiently connected to each other for processing.

In relation to the concept of commercialisation, datafication processes, enabling and enabled through digital data, facilitate value creation from these (see also Sadowski, 2019). Putting the concepts of datafication and commercialisation in relation to one another, therefore, allows the authors of the analysed literatures to interrogate what role in value creation processes is attributed to data. For example, Pybus et al. [39] alongside with other scholars [e.g. 10, 25, 28] critically attend to datafication as value extraction and argue that the value of data originates from the ability of companies, processing the data, to re-combine and relate different points to one another in a manifold of new ways [39, p. 3]. As Chen and Qiu [8] argue, however, human actors are extensively involved in this process of value creation: "the source of power for digital platforms – datafication – depends not only on algorithms, technology, and non-human natural resources, but also, more crucially and intensively, on humans themselves" (p. 287). They draw attention to datafication both as value extraction and exploitation of labour. Other authors explore how activists develop communities and services built on data infrastructures and how within these they resist commercialisation [20, 27]. For example, Kannengießer [23, p. 6] addresses commercialisation of data through the voices of data activists, who build awareness about the ways big tech companies extract value out of the data without giving any account of it to their users. With another example of data activist practices, Jackson and Kuehn [20, p. 421] notice how datafication, meaning here usage or sale of data stored by privately owned companies, alongside with venture capital, are the prevailing funding approaches in current economies. Analogous to this argument, Lehtiniemi [27, p.

635] explores how some businesses, such as personal data spaces, are not able to escape the pressure of datafication and instead imagine alternative strategies to redistribute the value of data processing, usage, and sale in different ways, granting their customers some more possibilities to decide, how exactly their data will be used and monetised. Not in the context of data activism but in a study of businesses, Beer [5, p. 30] also states that for many companies providing data-driven services to comply with the principals of datafication—fast-paced data processing, use, and sale—remains the only survival strategy.

Another perspective on the relations between the concepts of commercialisation and datafication offer Masiero and Das [32] in their analysis of anti-poverty programmes addressing the ways in that economic, material values (in this case material goods distributed through anti-poverty programs) are related to data.

> "The lens of datafication, largely used to interpret development in the domain of market and business intelligence, thus acquires a specific meaning when placed in the context of anti-poverty programmes. Such a meaning is centred on the twofold act of *recognising* beneficiaries, allowing to discriminate (for example through income data) the entitled from the non-entitled, and *assigning* the right entitlements to each of them" ([32], p.918, original emphasis).

In this sense, datafication processes reinforce existing inequalities in cases where both power over data and power over material goods is concentrated in the hands of the same actors. Similarly, Halkort [14] underscores how in the case of "datafication in the lives of Palestinians[, e]very attempt to measure and document their lives is read through the matrix of wider geopolitical interests and the historical experience of displacement" (p.322). This cluster of conceptualisations of datafication processes switches attention from the impact digital data make on individuals and societies to the political, historical, cultural, and economic processes in which data are being attributed to people and 'impact' attributed to data.

5.2.4 Re-situated conceptualisations of datafication

The overview of conceptualisations of datafication, related concepts, and terms provided above focuses primarily on those concepts the authors of the sampled literatures refer to as a starting point for their empirical investigations. In most publications, further, empirically driven understandings of datafication have been articulated that situate discussed definitions within specific empirical sites of practice examined by the authors. The remaining part of this section 5.2 attends to these empirical, re-situated understandings. Following Latour's (2004) argument about critique that needs to attend to concerns rather than facts, the understandings of datafication processes are clustered according to six concerns relevant for my analysis:

1) concerns about the implications, effects of datafication,
2) concerns about individuals, their identities, everyday practices, and subjectivities,
3) concerns about social justice and actionability[17],
4) concerns about power distribution,
5) accountability of various actors, and
6) concerns about "objectiveness" of datafication processes and digital data enabling these.

Some of these concerns directly respond to the conceptualisations discussed so far based on previous work in data studies. Others, such as concerns about accountability or "objectiveness", while being an integral part of some of the discussed conceptualisations, are only being put to the fore effectively in re-situated, empirical perspectives on datafication processes. In the following, the analysed academic articles are synthesised alongside these six concerns.

Within the sampled literatures, concerns about implications of datafication processes are described and situated differently, ranging from temporal and spatial transformations to

[17] The use of the term 'actionability' here is expanding on the concept data valency of actionability introduced by Fiore-Gartland & Neff (2015).

implications for democracy. Some scholars addressing implications of datafication emphasise different, new temporalities that datafication enacts [e.g. 7, 11, 15]. For example, Candido [7, p. 149] in her study of datafication in education, notices that data produced at schools and later used for planning and setting targets for upcoming years are temporally disconnected from people (educators and students). Duguay [11] provides another example, discussing how platforms give their users "sense that the news stories displayed are occurring live" (p. 27). Other authors underscore new spaces in which various actors acting upon data come together to perform or resist datafication processes [e.g. 2, 16, 18]. Finally, in some of the analysed articles, particularly possible negative implications of datafication processes are addressed, such as frictions within the societal domains in that it is performed [46, p. 462] or risks for democratic values [41, p. 10]. Thus, definitions of datafication processes concerned with their societal implications and situated in specific empirical sites of practices for which these implications are important, emphasise the relations between individual actors (educators, students, platform users, activists) among each other and to further, collective actors (schools, social media platforms, public authorities, and democratic societies). The authors consider how datafication processes reconfigure these relations between and among individuals and other kinds of actors.

Closely related to the examination of implications of datafication is an analytical focus on datafication processes in relation to the subjectivities and identities of individuals and the ways they configure/shape each other [1, 3 6, 13, 15, 25, 37]. For example, Barassi [3] emphasises how datafication is shaping citizenship in her analysis of 'datafied children':

> "[i]n order to understand the emergence of the datafied citizen, instead, we have much to gain if we focus on process, on the multiple ways in which individuals are being turned into datafied citizens, on the policies and political economic structures that make this datafication possible, and on the lived experience of this techno-historical transformation" (p.426).

Others take as a starting point in what Roberts-Holmes and Bradbury [43] notice in regard to datafication yielding changes in values and subjectivities in the educational domain [see 6, 37]. Baack [2] specifically attends to people's skills required to act within datafication processes, while e.g. Kennedy and Hill [25] attend to the affective side of lived experiences in datafied societies. The authors addressing datafication processes in relation to individuals' subjectivities are warning against reducing people to the 'datafied subjects' or 'data doubles', underscoring how such datafied representations of individuals not only oversimplify complex personalities and histories, but also possess normative power to define and categorise individuals "as fitting the norm or deviating from it" [6, p. 18]. Closely relating datafication processes to the changing subject and object positions people acquire through these, the authors explore datafication from the lens of the everyday: ordinary, routine activities, feelings, and perceptions through which these subject and object positions are continuously being renegotiated.

Authors attending to datafication from the perspective of social justice and actionability discuss inequalities as well as empowerment, participation, and practices of resistance performed by actors facing datafication processes in their lives. For example, Heeks & Shekhar [18] highlight different perspectives on datafication in international development: those attending to possible positive 'impacts', and those focusing on concerns related to these processes, the latter being about e.g. inequalities and exclusions. With their proposed conceptual model of data justice, the authors aim to provide an analytical tool for studying 'data systems' with regard to related "rights, structures, interests - in order to fully understand the implications of datafication" [18, p. 1007]. Dencik et al. [9] address datafication together with the commercialisation through the lens of data justice, arguing that the data-based economy drives not only privately owned, commercial companies, but also other actors, to join the race. Similar to the argument made by Masiero and Das [32], Dencik and colleagues [9] outline how forms of data-driven governance produce economic inequalities. Recounting not only negative, but also generative implications of datafication processes, other authors address new modes of participation and ways to engage with

data [7, 24]. Others, e.g. Baack [1], Kannengießer [23], Lehtiniemi and colleagues [27, 28] address various kinds of agency that activists and other stakeholders facing implications of datafication processes have and how they perform and/or resist datafication processes aiming to achieve more social justice. Pybus et al. [39] explore the possibilities to 'give back' and create access for people to their digital data and the ways such research interventions allow "to critically engage power-knowledge relations" (p. 8). Based on the discipline and research interest, some authors highlight how political actors become actionable and act upon datafication processes, while others primarily examine how actors such as activists or researchers alike exercise actionability by resisting, gaming 'data systems' and their extractive politics, and developing workarounds. Overall, concepts about datafication empirically situated through the concerns about justice and actionability underscore first, datafication as political processes, and second enactment of datafication as an ethical, moral, and normative issue. Can 'datafication' be productive if someone gives people their data 'back'? Should governments and public authorities have stakes in data-driven economy? While the authors propose practical, empirically grounded answers to these and similar questions, defining datafication processes as a concern about data justice and actionability they initiate discussions about the ways we as a society imagine just datafied futures and what we ought to do in order to reach these futures.

Scholars focusing on datafication in governance and policies address datafication as concerns about power and accountability. They explore and reconstruct various datafied regimes inscribed within and performed, enacted by these policies [e.g. 17, 36, 43]. Candido also argues that "[d]ata is used as a technology of government" (referring to works by Miller and Rose, e.g. 1990; as cited in [7], p. 129), therefore allowing for surveillance, monitoring, and control both on the state level and within schools [7, p. 151]. Similarly, Masiero and Das [32] refer to datafication through the idea of 'seeing like a state' and consider datafication "as an integral part of the policies […]" (p. 929). The vision they propose "frames datafied infrastructures as related to the political decisions behind them, which illuminates the dynamics of economic governance […] that such infrastructures advance" (ibid.). Takayama and Lingard [46, in contrast, showcase "the severely curtailed ways in which datafication has been enacted in the Japanese education system" (p. 462) and critique primary focus on negative implications of datafication common in Western academic discourses, specifically addressing education and educational governance (p. 465). Different authors are also concerned with the changing meanings of citizenship [e.g. 3, 26, 32, 41, 47]. These concerns by and large can be summarised by the following quote by Taylor and Richter [47] highlighting how citizenship is being defined based on digital data available to the states: "[t]he power to define knowledge is also related to that of defining citizenship: who counts in the datafied city is determined, to a great extent, by what is counted" (p. 722). Analytical attention to datafication in governance and politics, second, addresses concerns about (state) surveillance, dataveillance, and "shareveillance" (Birchall, 2016, p. 1; as cited by [14], p. 322) [see also 14, 26, 36]. For example, Lee [26] discusses perils of datafication by exploring Chinese social credit system that

> "institutionalizes dataveillance via datafication, and enacts social control upon its subjects. At another, how citizens respond to the system on a daily basis demonstrates adaptive self-surveillance, and how data-driven authoritarianism becomes deeply embedded and programmed into individuals" (p. 964).

Halkort [14, p. 322] also points to the involuntary conditions under which people become datafied subjects of state control. In relation to both surveillance, control, and accountability, Piattoeva [36] also emphasises the argument made by Halkort addressing mistrust arising from the use of technologies and data infrastructures. In a different example of voluntary data sharing, Jackson and Kuehn [20] discuss how privacy "becomes yet another social stratifier that disenfranchises those who often have the most to lose" (p. 424). Bayne at el. [4], outlining a picture of the future of education argue for new regulations that go beyond privacy, but protect against the categorisation of individuals based on surveillance technologies (p.104). While in the previous paragraph political

aspects of datafication processes have been discussed in relation to kinds of actions taken upon these processes, authors examining datafication as a concern about power and accountability illuminate how data politics are being negotiated both in research and practice and how data are being (arbitrary) attributed to people and things. Such understandings of datafication situated within distributed power relations of various—collective—actors such as governments, public authorities, and states, allow to raise questions about who holds positions to enable datafication processes, who is counted, and to what extent the voices of the latter are accounted for by the former.

Other scholars [e.g. 6, 7, 8, 42, 44, 51] address shifts in accountabilities enacted through datafication processes related to policies and governance across various societal domains, as the authors notice how both agency and power of such actors are distributed. Summarising this point Candido [7] addresses "the redistribution of agency across socio-technical networks, compounded by human and non-human actors" (p. 150), in this case e.g. algorithms for allocating schools' resources based on data. Bradbury and Roberts Holmes [6, 42, 44] further the argument by turning attention to the quality assessment of educational organisations based on their digital data and the related shifts in defining the most important aspects of education.

> "This shift demonstrates how datafication and schoolification both reproduce the other – literacy and maths are more easily translated into numbers, while their increased prominence in assessments encourages teachers to spend more time on these 'core' subjects." [6, p. 16].

Similarly, Winter [51, p. 57] addresses datafication with a focus on assessment data used for educational governance through references to concepts developed by Lingard (2011). Some other authors also put forward the need and the lack of regulation required to hold some actors accountable for the ways in that they perform and enable datafication processes [e.g. 8]. The 'distributedness' of power and accountability highlighted in this paragraph shines the light on non-human actors such as data and algorithms, their agency within datafication processes.

Finally, as mentioned earlier in this section, many authors of the sampled publications offer critique of the seminal definitions of datafication, addressing their insensitivities towards agency distribution and related shifts in power relations between those affected by datafication processes and those holding power to enable these processes [e.g. 10, 13, 25] and emphasising limitations of a 'technological' understanding of datafication processes [e.g. 5, 24, 29, 32, 47]. Such understandings of datafication are primarily concerned with the arguable claim of 'objectivity' of datafication processes. For example, Kelly and Noonan [24] warn against "the trap of treating "data" as an unproblematic "given" or as something that faithfully reflects an underlying organizational reality" (p.873). In the hindsight of this critique, the concept of datafication both at empirical sites of practice and in research, is addressed by some authors as an episteme with the help of which various actors understand the world [e.g. 7, 34, 36, 46, 47]. For example, Baack [2] discussing data journalism and civic tech activists, not only addresses their partnerships, but also the ways in that

> "[a]s a powerful emerging knowledge logic, datafication fundamentally affects how we collectively make sense of, and engage in our social worlds. Both data journalists and civic technologists aim to produce knowledge in the public interest and their entanglements affect the wider process of knowledge production and circulation in datafied publics" (p.2).

The quote illustrates how datafication can be understood as a way of 'seeing the world' through digital data and how practitioners from various societal domains such as activism and journalism, in this case, are affected by datafication processes in their practices of knowledge production. Translating this argument to academic knowledge production, application of the concept of datafication also implies certain view on the roles of digital data in social realities. An explicit reflection on the applied and developed conceptualisations of datafication and the ways these were developed empirically, may allow for a more empirically grounded understanding of datafied societies that is detached from the big tech industry discourses hailing datafication processes in form of data mining and data analytics (see Whittaker, 2021 for similar critical analysis in relation to 'AI'). In line with this argument some other authors of the sampled literatures expand on the

understanding of digital data or algorithms as assemblages, underscoring the socially embedded view on digital data and technologies [e.g. 17, 35, 45].

In sum, my analysis shows how, within the emerging field of datafication research, a number of overlapping concepts and related discourses co-exist. So, the concepts of datafication, quantification, commercialisation, and to some extent also concepts addressing digital data, such as digital traces or data assemblages, overlap in the sample of analysed publications. Putting these concepts in relation to each other allows to better understand the concept(s) of datafication in multiple ways. First, by drawing connections between datafication and established concepts such as quantification and trust in numbers, current datafied processes can be situated historically both at empirical sites of practice and in academic practices of knowledge production and discourses. Second, positioning datafication amidst other concepts allows to interrogate how datafication is connected to power issues. While for some researchers—using concepts such as commodification and commercialisation alongside with datafication—power is derived from economic value and value extraction mechanisms, for others power is rooted in the positions some actors hold, e.g. governments. Common to most is recognition of power acquired through the technological means that allow speedy processing of big amounts of data required and supporting the goals and values of actors holding that power. Finally, attention to the concepts such as data justice and data colonialism not only situate datafication processes historically, but also demonstrate the shortcomings of other conceptualisations and shed light on the leeway for agency and resistance power-less actors have when confronted with datafication processes.

Multiple authors of the sampled literatures draw on several definitions of datafication and provide their own, empirically situated understandings of datafication processes. Such empirical situating can be seen as critical reflection on previous scholarship and further theory-building about datafication processes. Moreover, my synthesis draws attention to a dual understanding of datafication as an empirical phenomenon and an epistemic, discursive lens used both by academics and practitioners (see also Kitchin, 2014b, p. 3). Particularly in theoretical discussions on the relation between datafication and quantification the latter understanding becomes especially tangible. As a way of "seeing, understanding and engaging with the world through data" [7, p. 129] datafication is about the process of translation (see Freeman, 2009 for various conceptualisations of the term) of the world into data and vice versa: by theorising and empirically investigating datafication processes, scholars also participate in this translation. Here, translation can be understood as work of relating various elements of datafication processes to one another, such as relations of affiliation and membership (such as citizenship), relations of attribution, and relations between the social reality as it and as it ought to be. In the following section I discuss in more detail based on my sample of academic articles, by which means and in which empirical domains, according to my sample of literature, datafication both as an empirical process and as an episteme render individuals and societies into data.

5.3 Datafication of what: identifying core elements of datafication processes

After situating various understandings of datafication processes conceptually, I explore empirical cases presented in the analysed literatures. The explorative keyword co-occurrence analysis presented in the section 5.1 of this chapter indicates, first, that datafication scholars conduct their research projects across various societal domains and with various actors. Second, the authors of the sampled publications offer a variety of answers to the question of what they are talking about when talking about empirical datafication processes. In order to cluster the literatures according to these two aspects, I examined research goals, questions, objects of study, and samples or case studies as articulated by the authors in their contributions. This section specifically addresses the actors and empirical sites of practice as elements of methods assemblages enacted in empirical research reported in the sampled publications.

5.3.1 Empirical sites of practice

Most authors address work practices of professionals in various societal domains, listed in Table 5-2: the authors either focus on professionals and laypeople, including their experiences with datafication or on policies and non-human rather than human actors.

Research within some of these domains requires from scholars particularly detailed ethical considerations, e.g. when working with some activist communities [e.g. 23]. Further, some authors also report how state regulation restricts possibilities of academic inquiry [e.g. 26, p. 957-958]. A number of scholars specifically attend to the roles of study participants in their research [e.g. 4, 39], aiming to involve studied people and communities in the research process not only as informants, but also enable them to make decisions regarding the research processes and their individual activities in line with participatory research approaches (discussed in more detail in the next section 5.4). For example, Bayne and colleagues [4, p. 95-96] extensively report about the roles of the student research assistants in the data collection and data analysis processes, underscoring how the students' perspectives merged into the analysis of a campus social media site YikYak. Finally, while empirically attending to various societal domains, authors of the sampled literatures also aim with their articles to provide insights, guidelines, offer critique, and outline possible futures not only relevant for the academic communities, but also for the study participants [e.g. 12, 14, 18].

Table 5-2 List of societal domains in which datafication processes were studied by the authors of sampled research articles

Experiences and practices of professionals	
journalism	[1, 2, 12, 38]
data activism	[1, 2, 9, 27, 28, 49]
education	[6, 7, 16, 33, 50, 51]
software development	[4, 5, 22, 31, 40]
politics/public administration	[8, 14, 17, 24, 41, 46, 47]
Lived experiences of laypeople	
self-tracking	[13, 14, 21]
social media	[29, 35, 39]
affect	[19, 25, 45]
information and communication technologies for development (ICT4D)	[14, 32]
Policies	
education	[36, 37]
beyond educational context	[26, 41]

Along the lines of the categorisation according to societal domains also lie the distinctions in the sample in regard to the means by which these domains, people, and things are being rendered into data and back. For example, some scholars, primarily attending to the practices of ordinary people or professionals affected by datafication processes, explore social media platforms and their algorithms [e.g. 4, 11, 29] or various kinds of software, digital 'tools', and applications and how these are used by these people [e.g. 1, 2, 14, 45]. So, Bayne and colleagues [4] ask in their research article about a disappeared (and re-activated at the time of writing[18]) campus social media YikYak, "what might be at stake in the loss of anonymity within university student communities in a datafied society?" (p. 92). In another contribution, Saariketo [45] questions "how avid users of ICT encounter self-monitoring data of their ICT use and what kind of possibilities for reflection are

[18] In 2021, YikYak has been re-activated: https://yikyak.com/the-yak-is-back (accessed 08.02.2022).

hereby enabled" (p. 126). In these articles, social media and other applications used by different groups of people can be understood as means through which the everyday lives of these people are being affected by ongoing datafication processes: for Bayne et al. [4] the changes in platform anonymity policies constitute a research case for studying ongoing datafication, while Saariketo [45] is interested in the ways people reflect on their datafied experiences.

Other examples highlight algorithms applied in different information systems and categorisations, metrics these algorithms include, as means constitutive to datafication processes. For example, Maasø and Hagen [30] ask

> "[h]ow are metrics of streaming usage influencing strategic choices and actions taken by stakeholders in the music business, and what are the relations between the metrics of streaming and the algorithmic affordances of music distribution in music-streaming services" (p. 19).

Lee [26] examines "how China uses a social credit system as part of its "data-driven authoritarianism" policy" at the same time underscoring the recursive relation between policies and scoring system as "datafication […] offer[s] the Chinese state a legitimate method of monitoring, surveilling and controlling citizens, businesses and society" (both p. 952). This research attends to the ways in which algorithms are ordering social phenomena, categorising, and labelling people and their activities, critically engaging with the politics of data.

Few authors interested in people affected by datafication processes, specifically examine the 'impact' of these processes on individual's and organisational practices, while policy documents and policy/governance initiatives can be understood as means by which manifold empirical realities are being translated into data and back again [e.g. 7, 33, 37]. This research is not only engaging with the politics of data, but also with political, economic, and social contexts in which data and technologies enabling datafication processes are situated, in which the use of these technologies, the meanings and representations of data are being negotiated and renegotiated. For example, in her article Neumann [33] explores how schools face policy changes. She explores how

> "macro-level policy changes affected the daily operation of an English secondary school and inquires as to how the expectations posed by accountability policies, the norm of data-driven education management, and the school's student grouping practice are interrelated" (p. 2).

Others, while investigating practices and imaginaries of technology providers, address data infrastructures, both developed for commercial purposes [e.g. 5], state-enabled [e.g. 24, 32], and civic, activist ones [e.g. 20, 27] or those developed and maintained by a combination of various actors [40, 47]. So, in their analysis of a water supply system, Taylor and Richter [47] aim to

> "examine the struggle over what is counted, and who is included, in the context of Bangalore's urban water system and the city authorities' engagement with data technologies to reorganize and document the system's functioning" (p. 722).

As the quote summarising the authors' study goal illustrates, the data infrastructures of the researched water supply system are an object to change and negotiation in the political practices of the city authorities. These data infrastructures encompass both water pipes, cables, and sensors measuring and controlling "water leakage" [47, p. 727]. At the same time, the data generated through the sensors and the ways these data are processed and used by the involved actors, are constitutive to the whole process of public water distribution. Within this research strain, a few authors specifically interrogate the data infrastructures' non-human, technological elements [e.g. 36]. For example, in an article addressing the role of surveillance cameras in Russian national education testing, Piattoeva [36] asks, "what difference video surveillance technology is expected to make to other agents' action" (p. 83). Similar to other examples of datafication scholarship discussed here, contributions viewing data infrastructures as means enabling datafication of various societal domains, underscore an understanding of datafication as recursive, social processes that not only affect social practices and social actors, but are being performed in socio-technical systems in which agency and power are distributed among humans and non-humans.

So far, in this section I discussed the societal domains focused on in empirical datafication scholarship, and what is seen by the scholars as means enabling datafication processes (and, at the same time, enabled by them) like software, algorithms, data infrastructures, and policy documents. My analysis shows that not only technologies are understood as means constitutive to datafication processes, but also policy documents and initiatives within which these documents are being negotiated and developed. It underscores the recursive, performative ways in which datafication processes are intertwined with social realities across a variety of societal domains.

5.3.2 Actors and how these are addressed in empirical studies on datafication

Interrogating the means of rendering social realities into data addressed in datafication scholarship, I already mentioned some of the 'units of analysis'—who or what are the authors attending to in their research. Policy documents, systems' interfaces, and other publicly available documentation presenting reports about planned or ongoing datafication processes (e.g. media coverage) can be seen as an example of such a unit of analysis within the sample [e.g. 5, 8, 11, 36, 37 in the sample]. These foreground imaginaries and inscriptions of datafication processes in information systems. Most of the authors of remaining contributions view practices and perceptions of either individuals living and working in different societal contexts or other, collective actors such as schools or public authorities acting in the face of datafication processes, as their units of analysis. Some other authors specifically address certain socio-technical systems such as data analytics in public services [e.g. 10, 24]. In the analysed contributions, most authors address these units of analysis as a part of their research goal and the sampling strategy or case study. In the remainder of this section 5.3 I synthesise the literatures to explore how datafication processes performed through software, algorithms, policies, and data infrastructures, can be reflected upon methodologically across the identified societal domains.

Within the discussed societal domains, the authors address various groups of people involved in, enacting, and acting within datafication processes. Table 5-3 summarises different kinds of human actors—laypeople and professionals—who have been relevant for empirical studies of authors of sampled publications. Finally, some datafication scholars primarily investigate social media from various perspectives, including their interfaces, companies providing such services, but also their users [4, 11, 29, 34, 35, 39].

These listings indicate a disbalance in the sampled literatures between studies focusing on 'ordinary people', their uses of various technologies enabling datafication processes, and everyday experiences with these on the one hand and a majority of studies investigating work practices of various professionals, both facing implications of datafication processes and enabling these due to their affiliations with technology providers or political, public actors on the other hand. Many of the analysed contributions focus professional activities within various societal domains such as e.g. journalism or education.

Table 5-3 List of actors addressed in the sampled studies on datafication and respective references

Laypeople	[e.g. 13, 14, 15, 19, 21, 25, 32, 35, 45]
Professionals	
Journalists	[e.g. 1, 2, 12, 38]
Activists	[e.g. 1, 2, 9, 23, 27, 28, 49]
Educators	[e.g. 6, 7, 16, 33, 50, 51]
Software and infrastructures providers	[e.g. 22, 31, 40]
Policy-makers or public servants	[e.g. 14, 17, 24, 41, 46, 47]

Further analysing this disbalance, in my reading of the sampled contributions, informed by a methods' performativity approach and theories of practice, I distinguished between studies with following three kinds of research goals:

- First, some scholars investigated *lived experiences and practices of individuals* [e.g. 2, 3, 7, 13, 14, 15, 19, 21, 25, 29, 30, 35, 39, 45, 51].
- Second, *research interests and goals aiming to understand organisational practices, the ways datafication processes are performed, renegotiated, and resisted within and by various commercial, public, or nongovernmental actors*, in some studies conveyed through individual representatives of the professional groups [e.g. 6, 7, 10, 12, 16, 17, 20, 22, 23, 24, 27, 28, 31, 38, 40, 41, 46, 47, 48, 50].
- Finally, another strain of research foregrounds *'impact' of datafication processes—their socio-political, socio-economic, and socio-cultural implications on bigger groups of people* [e.g. 4, 18, 27, 32, 33, 34, 36, 41, 42].

As these listings of sampled articles suggest, some researchers connect their research interests in collective perspectives and practices to individuals' experiences of being affected by datafication processes. So, Bradbury [6] in her analysis of yearly year education in Britain, asks "how schools and nurseries are subject to the demands of data, and how this produces data-driven teacher subjectivities and a 'data double (…) of each child" (p. 7). Providing an overview of several research projects, the author is interested in the ways schools and nurseries are affected by datafication processes manifesting in demands of data and how these processes ultimately relate to educators' subjectivities. Similarly, Macgilchrist [31] in a study of edtech (educational technology) providers also relates providers' perspectives to the experiences of individual employees in her research question: "how these individuals express a commitment to ethics and equity" (p.78). In an analysis of these perspectives, however, Macgilchrist [31] addresses software providers as a company and a coherent actor, rather than an individual person: "The first data story concerns a for-profit data-driven literacy platform. *It describes itself* as being built around collaborative and self-paced learning […]" (p. 79, emphasis added). At the same time, the author also highlights different roles various people occupy in edtech industry (e.g. CEO, funders, etc., see p. 81-82, see also [27] for similar examples). Such a 'humanised', anthropomorphic conceptualisation of collective actors, if not explicitly reflected in empirical datafication scholarship, may complicate an understanding of collective dynamics, practices, and negotiations required to perform datafication processes. In contrast to that Johanes and Thille [22] specifically portray individual education infrastructure builders in order to understand their personal positions and values: "[b]y capturing the voices of the infrastructure builders in their own words, we can better understand their inner world, the stories in which they see themselves as characters" (p.2971).

Many scholars are interested in organisational practices and perspectives on datafication processes. For example, Engebretsen et al. [12], exploring practices of data journalism in newsrooms, or Jackson and Kuehn [20], investigating an open source platform cooperative, attend to actors that I will call 'collective actors' as in opposition to research investigating practices and perceptions of individual people. Another example illustrating datafication scholars' interest in such collective perspectives provide Chen and Qiu [8] exploring Chinese transportation platform DiDi and asking how

> "does an app-based platform grow into a digital utility provider? What are DiDi's datafication strategies, its complex relationships with the Chinese government as an "infrastructural state", and its labor practices, especially regarding the intensive labor performed by its driver-workers? How does DiDi resemble and differ from Uber in the platformization of transport?" (p. 275).

An empirical investigation of a collective perspective, however, proved challenging for several authors. For example, Redden [41] summarises her experiences in building access to the public administrations involved in data-driven decision-making in Canada.

> "A combined approach of cold calling based on job title, snowball sampling and approaching senior level civil servants proved somewhat successful in gaining access. […] It was difficult to find out about the specifics of big data applications from interviews. It also proved difficult to find publications outlining government big data applications, a problem that exists across nations. For these reasons freedom of information requests were used to collect internal documents and communications about big data practices." [41, p. 5].

Such a reporting goes in line with organisation studies and information systems research, arguing that socio-technical information systems applied in organisations and public institutions are quite complex and manifold; a single person, therefore, is quite unlikely to have an overview over all relevant processes and practices (see e.g. Laudon & Laudon, 2020 for discussions on the relation between information systems' and organisational complexity). Other authors of the publications sampled for my synthesis, such as Kelly and Noonan [24] highlight research on datafication processes as "a complex and difficult process that relied on skill, intuition, and hard-earned sensitivity to the intricacies of the empirical context in which we were embedded" and an "outcome of careful and painstaking consideration of a range of different perspectives and sources of evidence" (p. 879). These quotes indicate not only what skills and resources researchers require to conduct empirical work, but also the interconnections and ties between various actors, the complexities of their positionings at the empirical sites of practice, and the ways all of these are constitutive to understanding datafication processes in relation to collective actors (in the latter case, Indian public service organisations). Other authors also mention further challenges of studying datafication processes such as restricted access to digital data (e.g. social media APIs, see [34 p. 11]), ethical challenges of anonymity and privacy [e.g. 23].

For empirical datafication scholarship, these challenges bring into question the ways of identifying and defining empirical datafication processes. First, restricted access to actors and resources required for research only allow to develop a limited understanding, especially in regard to the collective practices and perspectives on datafication processes. Second, due to such a restricted access, for some scholars, only individual actors or public documents are available. While the former provide valuable insights into personal stories of the study participants and collective practices, the latter often transmit a coherent public image of an actor, while internal negotiations, conflicts, and practices are written out. As my synthesis illustrates, some datafication scholars are specifically interested in such an individual perspective and personal stories [e.g. 22], while others (re)produce an anthropomorphic picture of organisations, for example software or infrastructure providers [e.g. 5, 27, 31].

My literature analysis indicates that while many authors identify datafication processes as being enabled, performed, and sometimes resisted by collective actors—commercial or public organisations, governments, nongovernmental organisations, and activist communities—these authors in many cases attend to individual actors representing only a part of such an organisation in their empirical research, rather than investigating the role of collectivity or relations of affiliation of these individuals to the collective actors in the development of datafication processes. Another strain of datafication research, at the same time, benefits from attention to the individual: scholars interested in lived experiences of people affected by datafication processes provide detailed insights in practices, agency, subject positions, and affectivities of these people. Furthermore, the analysis presented in this section illustrates how scholarly attention is directed at people positioned differently in regard to datafication processes: some study participants (e.g. 'ordinary' people) are primarily seen as being affected by datafication processes in their everyday or professional lives while others like technology providers or political actors are rather seen as holding power to negotiate and enable datafication processes. In the next section of this chapter, I continue

discussing these arguments and elaborate in more detail on the methods and techniques applied by datafication scholars to study datafication processes.

5.4 Matters and methods of concern in datafication research

After I recounted various elements of the methods assemblages: how the authors of the sampled academic articles view datafication processes differently, what societal domains they explore, and what exactly they do to examine datafication in their research, I finally turn to the methods—understood here as techniques of data collection and analysis—they apply to identify these concerns and study datafication processes. Together with the results of my literature synthesis in the previous sections of this chapter, these methods as techniques can be conceived of as elements of methods assemblages. In the subsequent, concluding section of this chapter, I construct and discuss methods assemblages in more detail. This section is structured as follows. First, I attend to the contributions in which qualitative methods such as interviews, sometimes accompanied by ethnographic observations, document analyses, or other techniques are reported. These contributions, in turn, are grouped alongside the disciplinary and thematic similarities as well as kinds of actors involved in empirical research identified in the previous section 5.3. Second, I discuss research in that document analysis and policy research served as central techniques for data collection [e.g. 5, 36, 37], highlighting the methodological challenges of access to some actors enabling datafication processes such as software designers or policy-makers. Third, studies in which digital and computational methods were applied as primary techniques of data collection and analysis are discussed [e.g. 3, 4, 11]. Finally, publications reporting findings of participatory projects applying varying techniques are addressed [e.g. 39]. As some authors present findings from multiple projects or applied a number of techniques for data collection and analysis, the studies in the following paragraphs are clustered according to the authors' descriptions of their research design in the analysed literatures and my reading of these. Importantly, following the reports on research designs provided in the empirical articles, I do not place specific attention here on methods of data analysis. The reasons for that are multiple: for example, not all analysed academic articles provide enough detail on the analytical techniques; in some publications, applied analytical techniques are also used to create research material in the first place. Throughout the section, concerns related to datafication processes raised by individual authors are discussed. This section 5.4 summarises such concerns and the main methodological issues identified in my analysis.

5.4.1 An overview of methods/techniques

Most of the sampled academic articles include reports about qualitative research projects, among these, authors of 15 articles discuss predominantly single technique of data collection such as interviews [9, 12, 13, 15, 16, 22, 30, 31, 38, 40], focus groups [35], content analyses [5, 36], digital methods such as walkthrough [11], and theory-based methods such as 'doppelganger as method' (Pierejilewsky) and 'countermapping' (Redden 2018). Over 20 further contributions applied multiple qualitative techniques, also including interviews, ethnographies, document analysis, and digital methods (e.g. digital ethnography, see [3]). This overview illustrates that the most applied technique to study datafication processes is conducting various kinds of interviews with individual people involved in datafication processes—those facing implications of datafication processes or those enabling these processes through their work practices and positions of relative power. Interviews provide insights into everyday practices and perceptions of datafication, into decision-making processes, dynamics of technology development, and regulation through the accounts of involved actors and these actors' individual as well as organisational values. At the same time, other aspects, inherent to the actors' practices and their tacit knowledge, such as kinds of technologies or data they operate with or data representations (e.g. visualisations) they encounter may be bracketed out and 'othered' in the interviews. For exploring these tacit knowledges, some scholars apply

multiple complementary qualitative techniques of data collection alongside with the interviews, for example document analyses, ethnographically inspired observations, digital and auto-ethnographies.

In addition, some scholars used prompts for interviews, for example asking their study participants to demonstrate their typical practices of self-monitoring [13] or by discussing during the interview data representations produced by the study participants [e.g. 40]. For example, for Ratner and colleagues [40] data visualisations as interview prompts allowed

> "to map important decisions and ambivalences when developing visualizations and to get accounts of how the designers learned about teachers' needs as well as how their work depends on wider institutional circumstances" (p.26).

In publications whose authors applied computational techniques together with qualitative interviews [e.g. 4, 49] as well as in studies reporting about findings from multiple projects, the mix of methods allowed an iterative, abductive research process for different packages of a single project or for separate research projects following one another. Some scholars applied interviews and surveys [e.g. 4, 6, 27, 48, 50]. The combinations of various methods also allowed to validate empirical findings in iterative cycles of data collection. Further, they served to increase trustworthiness of results (e.g. by cross-checking information from different sources, see e.g. [41]) and to include different voices [e.g. 3, 8]. In addition, some of methodological choices indicate different kinds of values that guided researchers in their studies such as empowerment of study participants [e.g. 39], providing comprehensive overview [e.g. 10], or empirically validating a conceptual framework [e.g. 18]. To analyse the interviews, most scholars applied various inductive methods such as thematic analysis [e.g. 9, 16, 35], while some turned to other analytical techniques such as rich points analysis [e.g. 31]. Not all scholars provided in their contributions a detailed methodological overview over applied techniques of data collection and analysis [e.g. 36]. In the following, I discuss in more detail what techniques datafication scholars apply in their empirical work and what concerns they raise in these publications.

5.4.2 Qualitative techniques for understanding laypeople

In their study on digital traces in self quantification, Gerhard and Hepp [13] conducted interviews "to focus the research on a limited number of self-trackers and thus gain a deeper insight into the contexts of self-tracking and self-quantification" (p.689). With that, personal (hi)stories of self-trackers and their practices of quantification were addressed. In addition, to explore what digital traces self-trackers encounter in their practices, "the participants showed and therefore made accessible to us the kinds of traces they produced" (ibid. p.698). The authors describe their study as "a reconstruction of users' everyday practices and their various contexts" [13, p.696]. Similarly, Hand and Gorea [15] in their analysis of self-quantification practices, also conducted interviews in which study participants were asked to demonstrate technologies they use for self-tracking. Parisi and Communello [35] also explore individuals' everyday experiences of datafication processes with particular focus on dating apps and "relational bubbles" –the implications of their algorithms on their study participants' app use. Janasik-Honkela [21], in turn, attends specifically to emotion tracking by analysing an emotion tracking app through related web pages and other materials such as YouTube videos, and interviews with the tracking app users, additionally exploring app use data of anonymous users (p.553). Empirical findings presented in the article primarily address the development of the tracking app and experiences of the app creator and app users. Recounting negotiations about how to describe and represent emotions in an app in the ways that would allow various users to recognise and track their own, particularly negative, emotions and how users encountered the outcomes of these negotiations in their tracking practices, the author raises a concern, among other, about "significant discords at the interface between the technology designers and the emotion professionals" (ibid., p.556) as well as app users at the same time highlighting a generative, productive aspects of emotion tracking.

Offering another perspective on datafication processes and emotions, Kennedy and Hill [25] engage with the various feelings that these processes evoke in people's everyday encounters with data and data visualisations. The authors "captured data about participants' encounters with visualisations in their everyday lives and about their engagement with specific visualisations circulating in the everyday" [25, p.836] through diary-keeping and focus groups with the same participants. By analysing feelings evoked through datafication processes, Kennedy and Hill [25] and Saariketo [45] switch attention in the academic discourses on data literacy by pointing to

> "the consequences of the ways in which people currently learn to relate to data through formal mathematical education – namely, it often results in a lack of confidence in numeracy skills. In our research, alongside such widespread lack of confidence in statistical literacy, we saw emotional relating to data, or feeling numbers" [25, p. 846].

Saariketo [45] created a research intervention with a goal of repurposing "the means of datafication to create possibilities for people to reflect on what it means in their daily lives" [45, p. 125]. The author, however, reports on a disconnect between the study goal to encourage critical reflection and the participants' motivations for taking part in the research project (see ibid., p. 129). So, study participants were curious about their app and media use. In the broader context of such datafication processes as data tracking and extraction, not only through specific tracking apps but also other media applications used in the everyday, this disconnect illustrates how actors affected by datafication processes act within these based on their personal interests and values, including affect. Further, Hill and colleagues [19] explore emotional implications of datafication processes through encounters with journalistic data visualisations. The authors analyse how newspaper readers encounter data visualisations that the authors "did not want […] to be defined as "good"" [19, p.339], also drawing attention to professional standards and practices of journalists creating such data visualisations.

Another example of research on implications of datafication processes on 'ordinary' people can be drawn from the contribution by Masiero & Das [32] who address Indian anti-poverty programmes, in particular its public distribution system and the related Unique Identification Project Adhaar. Conducting interviews with various actors involved in Indian public distribution system and attending to their lived experiences and various elements of the systems infrastructure (e.g. ration shops, biometric monitoring devices/scanners, etc.), the authors draw attention to legal, design-related, and informational forms of data injustice going along with datafication processes. Despite interviews being the main data collection technique, in their contribution the authors provide details on the historical development and goals of the anti-poverty programmes, unpacking datafied regimes inscribed in their infrastructures from the perspective of data justice and raising concerns about politics of these regimes [32, p. 930].

In sum, authors exploring how datafication processes affect 'ordinary' people by applying qualitative techniques of data collection primarily attend to the situations in that these people encounter digital data, their representations (such as visualisations), and technologies, infrastructures that make these accessible. Some authors are particularly interested in the emotional, affective part of these encounters and how digital data and datafication processes acquire value through affective relations [19, 21, 25, 45], while others explore people's lived experiences within and upon such encounters [e.g. 13, 35]. At the same time, some authors explicitly attend to technologies and data infrastructures that make encounters with digital data and their representations possible [e.g. 13, 32]. Masiero & Das [32] additionally offer a different approach, addressing data infrastructures as a part of historical development and across multiple technological actors.

5.4.3 Qualitative techniques for understanding professional practices

Other authors address professional practices of individuals and collectives in more detail, attending either to the implications of datafication processes for various professions [e.g. 30] or to the politics

of data these individuals and collectives negotiate and perform, enabling datafication processes further. For example, while Hill et al. [19] attended to newspaper readers and their encounters with data visualisations, others focus on journalistic—and in some cases also activist—professional practices. For example, Engebretsen et al. [12] attend to journalists' practices of data visualisation through interviews with journalists working in newsrooms in Norway, Sweden, and Denmark. Through their analysis, Engebretsen and colleagues recount the professional practices of journalists not only in creating the data visualisations themselves, but also the encounters with these visualisations, e.g. by educating the readers [12, p. 7]. Highlighting the positions of readers who encounter journalistic data visualisations 'on their own', the authors address concerns of alleged objectivity of such visualisations and data literacy required to understand these [see 12, p. 13]. Porlezza and Splendore [38] and Baack [1, 2] explore data journalists' work more broadly. In a case study about Italian data journalists, Porlezza and Splendore [38] conducted interviews according to a reconstructive approach that allows "exploring different facets of news processes" (p.1238). Reconstructive interviews provided detailed accounts of practices, reflections on personal (hi)stories of the study participants and addressed concerns about transparency and accountabilities of data journalists by underscoring "[t]he heterodox educational path, together with the need for specific skills in a highly competitive and professional sub-field" [38, p.1245]. While in the analysis the kinds of digital data or methods used by the journalists to collect and analyse these data, software, and instruments they used were left out, the authors discuss the interconnections between data journalism and politics shedding light on the journalists' pending between the implications of datafication processes on their professional practices and the necessity to translate these datafication processes to the journalists' audiences. Another perspective on data journalism provides Baack [1, 2], exploring both data journalist professional practices and data activist communities through qualitative interviews with participatory mapping [2] and content analysis [1]. The applied techniques of data collection allowed the author to examine the study participants' visions and accounts of practices, while particular kinds of data, software, and technologies used to operate with data were also included in the findings in one of the studies [2].

Other datafication scholars using (mostly) qualitative methods shed light on possible generative, productive implications of datafication processes exploring activist communities that resist inequalities, exclusions, and invisibilities brought forward through datafication processes [20, 23, 27, 49]. For example, Jackson and Kuehn [20], in their study of the platform cooperative Loomio.org attend to its resistances against commodification and datafication processes using a combination of discourse analysis and political economy (p.417). Exploring Loomio's infrastructures and relations to other actors, the authors underscore the tensions an open-source platform cooperative faces and a leeway of agency it and its users have. Similar approach takes Lehtiniemi [27] in his examination of personal data spaces (PDS) and their "intervention in the value accumulation model and the unilateral market operations of surveillance capitalism" (p.627). By applying qualitative interviews with various PDS developers combined with a policy questionnaire to understand the organisational dynamics of PDS providers and their infrastructures, the author discusses what role PDSs play in datafication processes in regard to agency and creation of (economic) value. By addressing PDSs as rather homogenous providers, e.g. "PDS do not only expect […]" [27, p.634], however, internal issues become bracketed out through the wording and a coherent picture form the outside is created. In another study focusing alternative datafication processes, Trevisan et al. [49] analyse multiple activist story banking initiatives through a mixed-method approach combining interviews with story banking professionals and an analysis of their LinkedIn data. The authors raise concerns about participatory practices offered through story banking among advocacy organisations, while they do not specifically focus on those who tell their stories, later archived in story banks. This latter point is common to the studies focusing on how professionals face implications of datafication processes in their work practices: when examining professionals' practices of enabling or resisting datafication processes, other relevant stakeholders

can be left out. Further, as the previous section 5.3 of this chapter showed, only some authors apply techniques that allow understanding collective perspectives (e.g. that of a newsroom or an activist community); rather, most authors conduct interviews with individuals across various organisations, while each individual recounts their perspectives on collective, e.g. organisational practices. In an exploratory ethnographic study with activist communities organising public CryptoParties in Germany, Kannengießer [23] provides an exemption by attending to the actors involved in CryptoParties, their "motives and aims" (p.2). Conducting virtual ethnography "at the online platform, especially focusing on the content advertising for the events" [23, p.4] together with participant observations and interviews at CryptoParties the author identifies these as communities and showcases their agency to empower 'ordinary people' and to resist "threat to privacy" (p.12).

Finally, Dencik and colleagues [9] in their study on political activism and anti-surveillance resistance report about social justice activists' experiences and resistance against data-driven surveillance (or the lack thereof). The authors introduce the notion of 'data justice' in an attempt to consolidate "those concerned with technology issues and those concerned with social justice issues" [9, p.10] and draw their attention to both social, economic, and political aspects of data-driven surveillance. The authors focus on activists' practices and perceptions required to resist surveillance, while the datafied regime enabling such surveillance was present in the article through the 'Snowden leaks'. In a further study of data-driven governance in the UK, Dencik and colleagues [10] explore perspectives of public sector and civil society on "data-driven forms of what we term 'citizen scoring'" [10, p. 1]. With a combination of semi-structured interviews, participation in workshops with relevant stakeholders, and Freedom of Information requests, the authors examined practices and concerns of the study participants in regard to data-driven governance. Dencik et al. [10] highlight how "little is known about the kinds of systems in place, how and where they are used, and what practitioners and stakeholders think about these developments" (p.19). The quote points to methodological challenges many scholars interested in datafication processes face such as lack of information about the technologies in place and difficulties of access to this information (thus, in this case, Freedom of Information requests were used to acquire missing information). In contrast to previously presented studies on activist communities [e.g. 1, 2, 20, 23, 27, 49], Heeks and Shekhar [18] specifically address the impact of the community mapping projects and datafication processes these enact in relation to marginalised people. In their empirical application of a data justice framework, the authors explore community mapping projects in Global South by conducting interviews with "with senior figures in the data intermediary organisations […] associated with each of the four projects" [18, p.997] combined with an analysis of evaluation documents from each project and secondary sources analysis. Although the authors encountered positive implications, they underscore limited agency of 'ordinary' community members in regard to the data.

A different example of an analysis of datafication processes and professionals' work provide Maasø and Hagen [30]. The authors conducted interviews with music industry professionals about their practices and the role of algorithms and metrics in these. The interviews "provided us with more insight into the various types and uses of data in the music industry, as well as a deeper understanding of how music metrics may have changed over time" [30, p.22]. With that, detailed accounts of the study participants' practices including particular datasets they were working with, and relations with other stakeholders (such as e.g. streaming services) were given. Attending to the professional encounters of study participants with metrics, the authors highlighted individual perspectives of the music industry professionals, while other actors such as music listeners and their listening practices are rather understood as 'data'. Thus, Maasø and Hagen [30] raise concerns about redistribution of agency between various collective, organisational and individual actors as well as non-humans such as metrics, as the following quote illustrates.

> "The true power of MSSs [music streaming services] resides in the way in which they have become central information hubs, with links and feedback loops to all of the other

> stakeholders in the music business – either directly, through the interfaces and algorithms they control, or indirectly, through data gathered from partners and intermediaries" [30, p. 29].

Other examples of research attending to actors facing implications of datafication processes can be found in the studies on datafication and education [e.g. 16, 33, 42, 51]. These studies not only explore educators' practices of enacting, acting within and coping with datafication processes, but also examine how educational policy regimes enable these datafication processes. For example, Hardy and Lewis [16] with their study on school performance data in Australia provide historical and political context of datafication processes enabled by the NAPLAN program in Australian educational domain. The authors' argument is similar to the one made by Maasø and Hagen's [30, p. 24] regarding music listeners understood as 'data' based on a number of characteristics such as their age, location, and the number of listeners/followers an artist has. Recounting concerns of datafication scholarship about the representations of users, particularly vulnerable individuals such as children, Hardy and Lewis [16] address the kinds of (in)visibilities enabled through educational measurement. The authors also highlight the generative possibilities of datafication processes in creating

> "spaces to be developed beyond more dominant logics and terrains of power, which helped students who participated in the programme to become 'visible' to one another (and their teacher) through the development of more substantive social and academic relations" [16, p. 246].

Despite these generative possibilities they underscore the role of datafication processes enacted through educational policy as these bring forth "pressures of comparison and performance, and the processes of data visualisation that make this possible" (ibid.) Roberts-Holmes and Bradbury [42, 43, 44] and Winter [51] also examine educators practices guided by particular educational standards in early years teaching in the UK, also attending to the lived experiences of educators, regulated by policies enabling datafication processes. The authors raise ethical concerns that follow datafication of education, summarised by Winter [51]:

> "[t]he regulatory curriculum system locks students and teachers into a totalizing technical–rational framework in such an all-embracing way that spaces for considering, deliberating about and acting on ethical responsibilities for and to others are elided, screened away, hidden, denied" (p. 70).

In another analysis of educators' practices of coping with datafication processes in the British educational domain, Neumann [33] explores how ability grouping in a secondary school is performed by conducting an "analysis of school documents and the school's statistical database" (p.5) combined with interviews with educators and focus groups with students. The inclusion of students' voices in the study sheds light on their experiences of competition and related feelings of insecurity. The analysis of school documents and educators' practices, at the same time, highlights the interrelation of these experiences with educational policies and their datafied regimes that "created the norm of public competition between schools based on exam performance" [33, p.19].

In sum, in studies addressing implications of datafication processes for professional practices, mostly individual perspectives of interviewed professionals are addressed. By developing research designs and sampling strategies that include multiple actors from one (or few) organisations/communities, a few authors specifically attend to the implications of datafication processes for collectives [e.g. 23]. In contrast to the analysis by Dencik and colleagues [9, 10] or studies of educators' professional practices [e.g. 16, 33, 42], scholars exploring professional and activist communities [1, 2, 12, 20, 23, 27, 30, 38, 49] primarily addresses the lived experiences of activists or professionals and their encounters with various digital data. The former, in turn, rather explore concerns about datafied regimes performed through policies and related issues of accountabilities, unequally distributed power relations, and their implications on (professional) practices of people in various societal domains. Exploration of lived experiences in such studies

underscores recursivity of datafication processes enacted in such datafied regimes. Particularly document analyses and policy research allowed scholars to attend to these concerns. Some of the reported projects attend primarily to the ways datafication processes are defined, described, and imagined in policies, while other projects attempt to examine their implications for stakeholders. While the focus on policies and policy-makers provides insights about relations between datafication processes and various societal domains, not always does it allow to examine implications of these processes for further stakeholders. So, only Neumann [33] included voices and perspectives of actors who are being datafied—in this case, children—in her analysis. Other scholars take the perspective of their study participants—professionals from various societal domains—in attending to further stakeholders as 'data', while also raising important concerns about their (in)visibilities, surveillance, and lack of agency.

5.4.4 Qualitative techniques for understanding enabling of datafication

While previous examples illustrated studies conducted with individuals and collectives facing implications of datafication processes, other scholars switch attention to actors who in their professional practices participate in enabling various datafication processes to some extent [22, 31, 40]. For example, Johannes and Thile [22] examining professionals building data infrastructures for higher education institutions worldwide, aimed to "not to describe a person in a photorealistic way, but rather to describe how that person constructs and interprets a world he/she inhabits, navigates and acts in" (p.2963). The authors attend to the particularities of the data infrastructures developed by the study participants: "By capturing the voices of the infrastructure builders in their own words, we can better understand their inner world, the stories in which they see themselves as characters" (ibid. p.2971). Producing such personal accounts in their analysis, Johanes and Thille [22] recount what values are relevant for educational infrastructure-builders and in what relation these values are with the inscriptions made in the data infrastructures. In another study on educational software, Macgilchrist [31] also addresses software providers' reflections on their own products, particularly in regard to educational equity. The interviews with representatives of various educational software providers were accompanied by an analysis of relevant legislations and policies. Although some of the interviews [see 31, p. 81-82] illustrate heterogeneity of the perspectives within the software providers' organisations, in the presented empirical vignettes these companies are addressed as homogenous actors with coherent values:

> "The first data story concerns a for-profit data-driven literacy platform. It describes itself as being built around collaborative and self-paced learning, enabling teachers to truly personalize instruction for individuals and small groups" [31, p.79].

In another study of actors enabling datafication processes in education, in this case, in Denmark, Ratner and colleagues [49] conducted interviews "with people who have been central in designing the scales and/or visualizations" (p.25) of educational data from Danish national testing. With that, the authors were able

> "to map important decisions and ambivalences when developing visualizations and to get accounts of how the designers learned about teachers' needs as well as how their work depends on wider institutional circumstances" (ibid., p.26).

As Ratner et al. [40] conducted empirical work in two organisations involved in producing data visualisations, the authors were able to explore values and concerns of the data workers, decision-making and negotiations of meanings of data and their visualisations, raising concerns about power distribution among human and non-human educational actors such as Ministries of education, teachers, and test data, as well as about different data politics guiding actors enabling datafication processes.

The three latter examples illustrated how software or infrastructure designers in their professional practices enable datafication processes in education. Another strain of educational research on datafication focuses on the roles of political actors in enabling these processes. So,

Candido [7] also attends to datafication processes in schools by conducting interviews and observations in Brazilian public schools in order to analyse quality assurance and evaluation (QAE) policies enactment discursively (p.133-134). Defining data as "a technology of government" (ibid., p.129) the author attends to various forms of educational quality assurance data (e.g. ratings and rankings) and raises concerns about the datafied regimes negotiated through and inscribed in educational policies. By applying multiple qualitative methods, the enactments of these policies in schools as organisations [7, p. 133-134] could be analysed. With her analysis Candido [7] aimed "to shift attention away from what is actually measured when identifying QAE policy enactments within schools in order to offer some clues regarding how datafication penetrates into schools" (p.152). In another study exploring data-driven educational governance based on the study of data-based school monitoring, Hartong and Förschler [17] combined interviews with monitoring agencies' experts in Germany and the USA with online and document analyses. The applied techniques focused not only on the accounts of practices given by interviewees, but also on various human and non-human elements of datafication processes, and particularly data infrastructures. Takayma and Lingard [46] with their study of Japanese educational domain also explore how educational authorities at different levels (from municipal to prefectural to national) enable datafication processes and are themselves affected by these. By contrasting a Japanese case study of standardised academic assessments to the Western examples, particularly those from the UK and the US, the authors show limits of existing academic discourses about datafication [46, p. 449]. Specifically, they elucidate the differences between Western and Japanese datafied regimes negotiated and performed through educational policies and governance: institutional constraints for data work and interoperability (ibid. p. 462), important roles of individual experts [46, p. 463], and tensions resulting from these.

Beyond the education research domain, some authors also explore how software developers and policy-makers enable datafication processes negotiating and performing datafied regimes through their products or policy documents. For example, Chen and Qiu [8] report about another study aimed at identifying regimes enabling datafication processes in the context of "platformization of transport services in urban China" (p.274). With their examination of "the secondary materials collected from government, corporate, and news documents" accompanied by "the driver's insights and our [the authors'] knowledge of their perspective" [8, both p. 275] the authors contextualise datafication processes in China' transport services in the broader socioeconomic processes. Instead of attending to the technological elements of transport services platforms and data infrastructures, Chen and Qiu [8] trace the dynamics of these infrastructures' development and highlight their societal implications such as labour required to maintain these services (p.285). Lee [26] also engages with datafication processes in China with the help of multiple qualitative techniques such as scenario-based story completion method, analysis of policy, online discussions, and media reports. Exploring the Chinese social credit system, Lee [26] elucidates how datafication processes enable dataveillance and what leeway for agency citizens have (p. 964), also acknowledging challenges of conducting empirical study on the chosen topic including state censorship (see e.g. p. 957). Further attending to the social credit system "through the concept of public goods, technology, privacy and collective punishment" [26, p. 964], the author traces infrastructures further enabling dataveillance in China.

Other authors attend to the actors and policies enabling datafication processes in India [24, 47]. For example, Taylor & Richter [47] conducted interviews with various actors "involved in water provision and the implementation of new digital data and related technologies" (p.723) such as companies providing infrastructures and activist associations. By attending to the accounts of such actors, the authors follow technological infrastructures enabling datafication of water supply, movements of digital data across these infrastructures, and its implication on citizenship. With their analysis, Taylor and Richter [47] raise concerns about commodification of such goods as water reinforced by datafication processes, related issues of citizenship, and to accountabilities of

infrastructure providers. Similarly, Kelly and Noonan [24], applied multiple qualitative methods to examine and trace the developments of datafication processes in Indian public health services over a longer period (the project was ongoing since 2007 at the time of publication in 2017) (see p.878). The applied qualitative techniques (ethnographically inspired interviews, observations, document analyses) and attention to broader societal and political contexts (by examining policy documents and political actors) enabled to flesh out a perspective on datafication processes and data infrastructures enabling these in their 'becoming'. While in their study the authors report about datafication processes enabled by public health data infrastructures, the technological aspects of these infrastructures as well as the data moving across these are not in focus of the contribution. Instead, the publication by Kelly and Noonan [24] alongside with other sampled articles [e.g. 6, 44] illustrate another concern common for datafication scholarship. This concern addresses tracing the dynamics of policies and governance regimes mutually enabling and enabled by various datafication processes.

In a decolonial, posthumanist study of data sharing in Palestinian refugee camps in Lebanon, Halkort [14] offers another perspective on technological and data infrastructures enabling datafication of migration by attending to "the nonlinear transition of lived and embodied knowledge into and out of data" (p.317). In the study, historical analysis was combined with long-term observations, interviews, and focus-groups with various actors in refugee camps (both refugees and representatives of major involved intergovernmental organisations). With her analysis of both lived experiences of refugees and the "economy of data sharing" [14, p.323], the author addresses concerns about value extraction enabled by datafication of migration, agency available to refugees, and politics of digital data, also acknowledging ethical issues of research in refugee camps [see 14, p. 322-323]. An example of research on datafication processes that is more explicitly focused on regimes of governance provides Redden [41] using theory-inspired methods. In her analysis, she applies the technique of "counter-mapping [that] can be used to overcome power hierarchies and as a means to pose alternative imaginaries that challenge or complement standard representations" (p.4) alongside with semi-structured, qualitative interviews, documents analysis, and Freedom of Information requests. With the combination of these qualitative methods, the author attends to discourses common to data studies that were missing in the discourses of the policy-makers. This contribution addresses in detail the challenges of governance research, also acknowledging the heterogeneity of the ministerial departments responsible for datafication of public services in Canada [41].

Overall, the literatures addressing professional practices, values, accountabilities of actors enabling datafication processes, and related data politics also argue that software and infrastructure designers as well as policy-makers are not only aware, but also reflexive about their power positions and their roles as enablers of datafication processes [see 17]. The publications reporting on empirical studies with different technology providers or designers illustrate the challenge of studying actors enabling datafication processes such as access to study participants and materials they use in their professional practices [7, 8, 22, 31, 40, 41]. In the presented qualitative research on actors enabling datafication processes, close attention is paid to policies, how these are negotiated, and performed, while not all authors provide examples or, even less so, discuss particular kinds of datasets, data visualisations, or technologies such as software and data infrastructures that are central for enabling datafication processes [see 7, 17, 40 for exemptions]. Studies focused on actors enabling datafication processes, while highlighting the important roles of e.g. public authorities or software providers, only sometimes explore these as collectives of various heterogenous actors in order to disentangle how heterogenous values, interests, and related frictions are represented in the negotiations of datafied regimes and, ultimately, intertwined with individual enactment of datafication processes (some of the exemptions are e.g. [17, 40, 41, 46]. Finally, some of the authors specifically attend to the technological and data infrastructures that enable datafication processes and through which these processes are enacted [e.g. 47]. In their contributions, the authors also

illustrate various dynamics of these infrastructures such as their historical development over time [e.g. 14], data movement across the infrastructures [e.g. 8, 47], and ongoing negotiations of data representations enacted in these infrastructures or software [e.g. 22, 40].

5.4.5 Document analysis, policy research, and matters of concern they raise

While scholars whose work has been discussed in the last few paragraphs were able to gain access to actors enabling datafication processes such as political and governance actors, technology providers, and infrastructures builders, some of the authors explicitly address challenges of such empirical work and combine interviewing techniques with document or policy analyses [e.g. 41]. Other scholars, therefore, primarily applied methods such as document studies and content analyses in order to examine documents produced by software providers or policy-makers [5, 36, 37]. Other techniques such as close readings of technology providers' websites and their marketing texts also allow understanding of datafied regimes and values inscribed in the software. For example, Beer [5], in his analysis of data analytics companies identifies 'speed' as one of the core values of that industry. The author describes only text passages including 'speedy' keywords, bracketing out the extent to which 'speed' is addressed in the data analytics companies' texts and other keywords. At the same time, close attention to one aspect of the data analytics industry's datafied regimes allows the scholar to contextualise it in the context of neoliberal datafied economies [see 5]. Other scholars use techniques from policy research to examine how policy documents enable datafication processes and negotiations about what elements of these processes or their implications do they include. For example, Piattoeva [36] in her analysis of Russian national testing examined "education policy documents, newspaper articles and various internet sources in order to explore the recent history of video surveillance as a technology involved in administering standardised testing in Russia" (p.83). In this study, Russian regimes of surveillance are situated in broader political, governance contexts while accounts of practices of people physically involved into the examination practices and their related values or emotions were outside of the scope of the study. Rather, the author raises concerns around accountabilities in datafied education governance and policy, reflecting on the

> "profound mistrust of all human agents associated with the data assembly-line at its various stages, and mistrust of the human paves the way to firm faith in the capacity of surveillance devices to achieve objectivity" [36, p.94].

Another example of datafication processes enabled through educational policies and governance in the UK provides Pierlejewski [37] with her theory-based approach called "'doppelganger as method' [that] utilises analytical devices from literary criticism, psychoanalysis, Foucauldian analysis, digital sociology, education studies and film studies" (p.3). For Pierlejewski, the method can be explained as follows: "[t]he role of the doppelganger in policy is not overt. It is hidden from view in the unsaid and the implied" [37, p.6]. Through a close reading of one British education policy text, negotiations of 'good' data, the subject and object positions of schoolchildren and educators are analysed. As a published policy document is interrogated, the decision-making processes that led to the development of the documents cannot be accessed. Rather, as the author argues in the conclusion to her contribution, the doppelganger method allowed to examine the changing subject and object positions of actors in datafied education: "[t]he impact of datafication on education goes far beyond changes to the curriculum and pedagogy; it creates a different kind of subjectivity in the cyborg-self" [37, p. 10]. In sum, authors primarily conducting document analyses or policy research, explore datafied regimes enacted in these documents, reconstructing the kinds of data representations negotiated by political actors or technology, data infrastructures providers. As the analysis of the latter three publications shows, some scholars raise concerns about power and agency distribution between various human and non-human actors, their accountabilities, and goals as juxtaposed to the positions and values of actors ('ordinary' people) potentially affected by the interrogated datafication processes.

5.4.6 Digital and computational techniques

While some of the analysed contributions focus on the document and policy analysis, other scholars apply digital and computational methods that also allow to access information about datafication processes when relevant actors are inaccessible. For example, in her article, Duguay [11] argues that walkthrough as a method provides scholars with opportunities to study commercial platforms otherwise inaccessible for researchers. The walkthrough method enables datafication scholars to examine "features, interfaces, icons, text and symbols" [11, p. 23]. At the same time, a walkthrough provides a rather static picture—a snippet—of, sometimes highly personalised, graphic user interface. In a close analysis of the social media sites news user interfaces, walkthrough turns researchers' attention to the "visions and anticipated uses of their news functionalities" [11, p. 23]. Additional analysis of the sites' platform guidelines and policies together with the walkthrough allow speculations about the software affordances and datafied regimes inscribed in the software and its interface. Another example of the use of digital methods alongside with other, qualitative techniques provides Barassi [3] with a combination of auto-ethnography, digital ethnography, and interviews applied to study the role of datafication processes in parenting and childhoods in the UK and the USA. With the help of said techniques the author provided detailed first-hand accounts of parents' everyday practices, kinds of data they encounter, the manifold of actors acquiring access to these data, and the relevant data protection policies and regulations [see 3]. In contrast to the walkthrough conducted by Duguay (2018) that examines platform affordances and regimes inscribed in social media sites' news sections, Barassi [3] primarily attends to the lived experiences of individuals and families affected by datafication processes.

Studies in which various research techniques were applied within one or multiple projects allowed tracing dynamics of the observed datafication processes. Some authors such as Bayne and colleagues [4] specifically attend to the dynamics of technological infrastructures, including required financial resources, as they outline the history of the rise and decline of a social media site. By applying different qualitative (interviews, observations), quantitative (survey), and computational (topic modelling) techniques, researchers could additionally examine YikYak users' views on its development over time. Bayne et al. [4] provide a detailed reflection on their methodological choices, as they discuss how their iterative research design allowed to provide trustworthy and robust documentation of the development and decline of YikYak as a university campus social media. Papakyriakopoulos and colleagues [34] applied computational methods (LDA topic modelling) combined with other qualitative techniques (p. 4) in their study about ways to implement microtargeting for political campaigning in German social media. With an examination of digital traces left by Facebook users, the authors attempt to identify micro groups of users for potential targeting and discuss ethical and political implications of such practices, raising concerns about users' privacy and accountabilities of big tech companies such as Facebook [34, p. 11]. While the authors' main goal was to examine the possibility of microtargeting in German political context, Papakyriakopoulos et al. [34] also address the socio-political implications of microtargeting and social media's infrastructures allowing these for political and electoral behaviour. In sum, authors applying digital and computational techniques of data collection and analysis specifically attended to the issues of access to such data and related research ethics. While Duguay [11] primarily reconstructs datafied regimes as performed through social media new interfaces, Bayne et al. and [4] and Papakyriakopoulos and colleagues [34] attend to the data infrastructures and their social, political implications on various actors. Barassi [3] with a research design combining digital ethnography with other qualitative methods offers another perspective on the use of digital techniques and explores lived experiences of individuals affected by datafication processes.

5.4.7 Participatory research on datafication

Besides research designs including qualitative and digital, computational techniques, some authors conducted participatory research projects. Studies that are understood here as reporting participatory research projects can be described with the following quote by Leurs [29], illustrating how study participants "also directed the course of the study as they became co-researchers of their own digital practices" (p. 131). Particularly two articles [29, 39] also applied participatory studies to give voice to young people in the discourses about their increasingly datafied lives. Pybus and colleagues [39] were particularly concerned with the issue of data literacy: "[i]n short, we did not simply want to use our subjects as data producers, instead we wanted to imagine what a data literate subject might look like" (p.4). Scholars who applied participatory methods in their studies used a variety of data collection techniques: participant observations [28], interviews accompanied by digital methods to produce data visualisations [29], and focus groups also accompanied by digital methods and hackathons for participatory analysis of digitally gathered research materials [39]. In the two latter studies, digital methods were used to provide the study participants—and co-researchers—with some input about the datafication processes relevant to their everyday lives such as, respectively, social media and app use. While Leurs [29] applied participatory methods to elicit emotions from social media use by learning personal stories of the study participants, Pybus et al. [39] also conducted hackathons and created various prototypes to analyse data gathered by the scholars together with their co-researchers. Additionally, Leurs [29] reflects on concerns articulated by study participants in regard to the application of research software for creating visualisations of Facebook friendship networks as well as on his own position as researcher and its performative role for the reported study (p.145). Pybus et al. [39], in turn, reflect on "[t]he challenge with our methodological approach [that] lies in its demand for specific technological competencies to manipulate the data that we were able to provide" (p.7). In contrast, Lehtiniemi and Ruckenstein [28], working with data activists, recount in their article their own roles as researchers and their engagement in the activists' practices. They reflect on the discomfort and different imaginaries they and their activist partners had. The long-term engagement with the community of data activists enabled the authors to change their positions [28, p.4], their research focus, emotional connections to the project and the activists, and observe the development of the data initiative. In this study, however, the activist community is addressed as a whole, without particular focus on the ways the community and its collective practices were organised, including required financial resources or technical infrastructures [see 28]. Rather, the authors discuss activists' goals and values, relevant stakeholders, including other activist communities, and policies regulating their work. The research focus is placed on the activities performed together by researchers and activists [see 28]. Overall, my analysis illustrates how participatory research provided the authors of the three articles presented here with spaces, tools, and vocabulary for explicitly reflecting their positions at the empirical site of practice within the communities and groups of study participants. Close engagement with the lived experiences of the study participants—co-researchers—allowed to put focus on different aspects of their datafied lives: emotions, everyday, and professional activities, while various data representations (visualisations) were used as prompts to elicit information about these experiences.

In this section 5.4 I elaborated on various techniques of empirical data collection and analysis applied by datafication scholars in their publications. My synthesis shows that while most authors presented findings of qualitative projects combining multiple techniques, the units of analysis they address and the matters of concern they raise in their empirical investigations do not equally follow the well-elaborated distinctions between different research methods. Rather, different demarcation criteria can be identified, that assemble methods, units of analysis, and concerns. So, as elaborated in the previous section 5.3 *the degree of collectivity of actors* addressed in empirical datafication scholarship and p*ositionings of these actors on the continuum of 'users' affected by datafication processes and 'producers'*, holding some agency and power to enable these processes, become

central. Specific research methods of data collection and analysis, in turn, are applied in accordance with the research questions and in hindsight of practical issues such as access to certain human actors, technologies, documentation, and digital data. Mapping and reflecting on the developing field of datafication research methodologically cannot be done in alongside clear and neat distinctions between various research methods, as often presented in methodological textbooks. Instead, the datafication research in practice is complex, open to creative and innovative methodological approaches and empirical interventions. My analysis therefore indicates that for datafication scholarship to attend to its most pressing questions about the relations between data, society, and knowledge, a different methodological vocabulary that is in situ sensitive towards multiplicities of conceptual, empirical, methodological, and epistemological perspectives on the term 'datafication' is required. The literature synthesis presented in this chapter provides a foundation for such a vocabulary and builds a heuristic for methodological reflection. Before unpacking this argument in the concluding section of this chapter, I summarise the current development stage of the field of datafication research: data studies.

5.5 Assembling data studies

The literature analysis discussed in this chapter makes several contributions to the goal of my thesis in mapping current empirical datafication research, addressed through my sample of publications, in regard to methods assemblages enacted and conceptualisations of datafication developed there. First, my literature analysis situates empirical research on datafication presented in the analysed academic articles in the emerging field of the data studies. Second, publications analysed in this chapter are discussed in relation to the elements of the methods assemblages introduced in chapter 3 of my thesis: researchers and their positionings, the researched actors, the empirical sites of practice in which datafication processes are being identified and studied, and the research processes, practices, and techniques themselves. While the literature analysis provided here lays a foundation for identifying methods assemblages enacted in current empirical research on datafication, as I discuss further in this concluding section, analysing texts of published research has its limitations if the object of inquiry is empirical practice. In the next chapter 6, therefore, the synthesis that I began presenting here is complemented by interviews with the authors of these texts and their reflection of their own research designs, practices, and positionings.

Before I continue with the summary of findings from chapter 5, the role of data studies as an emerging field in the narrative of this chapter needs to be clarified. I have shown both in chapters 2 and in more detail in chapter 5 that examples of research projects addressed here as 'data studies' can be encountered in different disciplines and domains. Taking that as a starting point, this chapter took first steps in investigating the relation between data studies and other research domains, attending to the ways in which datafication-related concepts and terms travel and become associated across these domains. The findings presented here indicate that data studies, while cutting across various empirical and theoretical fields, have their own, shared, but distinctive body of work. In the analysed research articles, different concepts, terms, and matters of concern such as standardisation, commodification, accountability, surveillance, and other have been addressed both conceptually and empirically. Despite the heterogenous origins of these concepts, terms, and matters of concern, their varying attention to technological, political, economic, or societal issues, my analysis shows how these are being brought together and related to each other through 'datafication'. This regards not only research domains present in my sample of literature, but also other research areas not covered in the sample as findings from the section 5.1 indicate. An example is an article not included in the sample and addressing labour of Amazon workers from the perspective of theories of work (Delfanti, 2021). Delfanti provides an analysis in which datafication is not seen as an 'external' process to the work and labour practices, but rather is considered a part of it. Consider the following quote.

> "While scholars such as Shoshana Zuboff (2019) and Jathan Sadowski (2019) conceptualize surveillance capitalism as based on the "expropriation" or "extraction" of data from passive users, they tend to overlook the active role of the women and men who perform the processes that are datafied and incorporated into machinery, as well as the forces that push them to collaborate with data collection." (Delfanti, 2021, p. 40).

For a methodological inquiry like my own, such a contribution poses a challenge and highlights methodological limitations of the research design chosen for my thesis. Further, analysis of references to further academic publications suggests that the body of work distinctive to data studies cuts across more research fields and disciplines than those covered through my sample, including also health, human geography, computer science. My literature analysis, therefore, indicates how data studies are assembling as an emerging research field, associating and bringing into discussion scholars and concepts with heterogenous foci. I focused here on literature and only briefly attended to other relevant aspects of a field such as key institutional and social actors or the ways in which relations in the field are organised. Instead, following long-standing lines of research (such as e.g. Knorr-Cetina, 2002; Latour & Woolgar, 1986), in the next chapter 6 I focus on specific practices and reflections with which methods assemblages of datafication research are enacted empirically.

Within this developing field of data studies, there are multiple overlapping academic discourses (e.g. on quantification, commodification, digital traces, etc.), while an approach to the concept of datafication as an epistemic one ('seeing the world through data') is common to all. The first section of this chapter illustrated that the qualitative synthesis of the sampled literatures presented in this chapter can serve for mapping methods assemblages of data studies. On the one hand, the sampled publications reflect heterogeneity of research about datafication both in regard to onto-epistemological, theoretical, methodological, and thematic approaches within social sciences and in regard to geographic regions of studies. On the other hand, the bibliographic coupling and co-citation analyses I conducted indicate shared theoretical background on datafication processes in various societal domains. Among the concepts central or relevant for understanding datafication are approaches highlighting the technological elements of datafication processes (e.g. digital traces and data infrastructures), academic discourses providing a big scale, historically embedded view on datafication processes in relation to politics and economy (e.g. data assemblages, commodification, data colonialism), and theoretical frameworks that position datafication as a continuation of other social processes (e.g. quantification, mediatisation). The authors of the sampled literatures expand on these concepts and discourses in their empirical work in order to situate their understandings of datafication in observations from empirical sites of practice (section 5.2).

In section 5.3 of my analysis of sampled research articles, I discussed in detail some of the core elements of the methods assemblages such as empirical sites of practice where datafication processes are being studied and actors addressed in the analysed publications. The ways in which these cites of practice were chosen and the actors addressed are related to the kinds of knowledges sought by datafication scholars. I identified three different kinds of knowledges:

- Socio-political implications of datafication processes;
- Negotiating data representations;
- Lived experiences of actors involved into datafication processes.

To develop situated understandings of the term datafication according to the kinds of knowledges they seek, scholars attend to various means by which datafication processes are enacted and social realities rendered to data and back such as software, algorithms, and data, data infrastructures, policy documents and projects. Empirically, datafication scholars are interested in the ways that 1) individuals and collectives such as communities, organisations, and public administrations see, negotiate, and enact their lived realities in the face of datafication processes and 2) these individuals

or collectives are positioned within datafication processes (affected by these or having agency to enable or resist these). Finally, in section 5.4 I outlined the manifold of research techniques applied by datafication scholars and concerns either raised and explored in detail or othered, absent in the sampled publications were identified. In hindsight of the methodological heterogeneity and the identified continuums in understanding datafication processes empirically—extent of collectivity of addressed actors and their positioning in the processes of datafication—I argue here that current methodological vocabulary is not sufficient to reflect on the developments in the field of datafication research and the ways various concepts of datafication are developed in this field.

Therefore, I suggest the two identified continuums as a heuristic for reflecting on datafication research and data studies methodologically. First, in regard to the extent of collectivity of actors addressed empirically, and second, in regard to the positioning of these actors in datafication processes. Figure 5-5 illustrates these two methodologically relevant continuums. Depending on the societal domains, study focus, and an empirical site of practice including means by which datafication processes are being enacted, datafication processes currently examined in social sciences can be identified differently within these continuums.

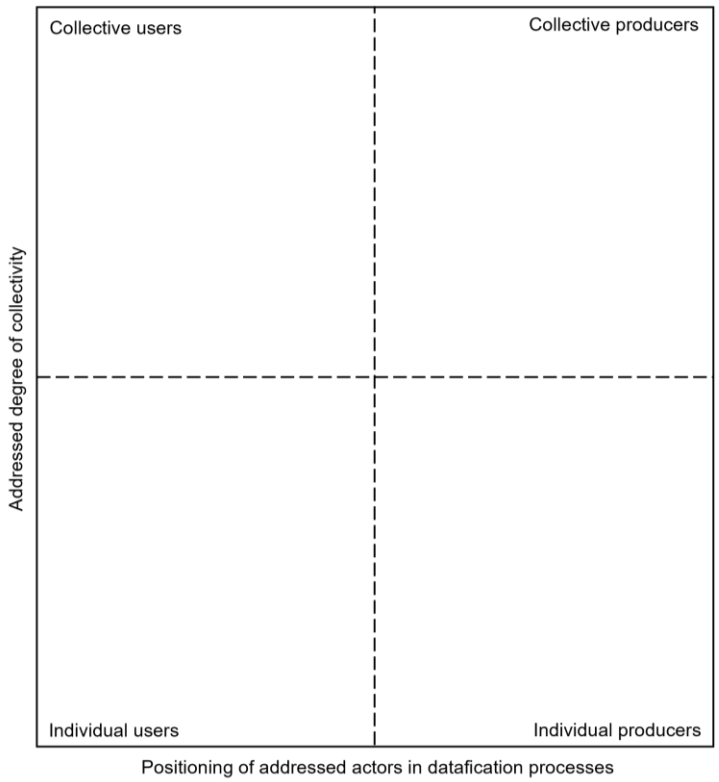

Figure 5-5 A heuristic for mapping research on datafication

In regard to the extent of collectivity addressed empirically, data collection through interviews with individual actors within such collectives rather underscores the heterogeneities within these collectives, possible differences in values and motivations of different members of the collective, and related conflicts than it provides a comprehensive perspective of the collective. It can be concluded, thus, that while a manifold of conceptual contributions to research on datafication, as presented in chapter 2 of this thesis, hint to the 'big scale', societal implications of datafication processes, empirical studies primarily offer individual perspectives on these processes. It brings into question the extent to which individuals' experiences and practices allow to understand datafication processes better or rather convey a partial view that needs to be acknowledged explicitly in reporting of the results. The analysis of the sampled academic articles shows that particularly document analyses and policy research allow scholars to answer questions about values and imaginaries of datafication processes for collective actors. Both individuals and

collectives are addressed by datafication scholars as either those affected by datafication processes who can be seen as users of socio-technical systems, through which these processes are being performed, or those in position to enable, negotiate, or resist these processes. As multiple contributions, particularly those addressing activist communities, illustrate, actors affected by datafication processes can exercise their agency to some extent and sometimes resist datafication processes. At the same time, collective actors such as businesses willing to continue their economically stable existence cannot fully resist datafication and commercialisation processes. This second continuum, therefore, can be defined as the positioning of the addressed actors in datafication processes between the poles of 'users' and 'producers'. Together, these two categories build a heuristic and vocabulary for developing and examining methods assemblages in more detail based on expert interviews with datafication scholars (figure 5-5).

In the quote referenced in the beginning of this chapter Latour (2004) argues that analytical work is bound to reduce complexity to present an argument. Even if accompanying datafication as a 'scientific object' back to its 'gathering' weakens datafication processes and research on datafication, my aim with the developed heuristic (figure 5-5) is to introduce methodological vocabulary which advances sensitivity to the multiplicities and complexities of empirical datafication processes. Initially, I analysed the sampled articles based on different kinds of methods and techniques the authors applied. My analysis showed similarities across these methods which go beyond specific techniques of data collection and analysis. Rather, the mapping of such methods and techniques presented in the section 5.4 suggests that datafication scholars conduct research that is driven by research questions, combining various qualitative, quantitative, computational, and other, new methods to answer these questions and achieve set study goals. The heuristic I started to develop with my literature analysis is sensitive to the central aspects of datafication processes such as kinds of actors whose perspectives and activities are investigated empirically, their positionings in the datafication processes, and the means by which these actors and other stakeholders perform these processes, as discussed in the next chapter 6. In this chapter of my thesis, I also argue that various kinds of knowledges sought by researchers conducting research on datafication and the matters of concern they develop in respect to these sought knowledges are relevant for identifying methods assemblages enacted in empirical datafication research.

The heuristic developed in this chapter provides space for assembling various elements of empirical research design in datafication scholarship, encompassing theoretical, epistemological, methodological, and practical elements of research practices and academic knowledge production. To reconstruct the methods assemblages actually enacted in empirical datafication research, however, it does not suffice to simply put the research methods/techniques on such a heuristic. Rather, it requires practice of reflection to trace the connections between specific research procedures and other elements of a methods assemblage: the scholars' positionings in academia, their theoretical and epistemological assumptions, the actors they bring together through their empirical studies and the addressed empirical sites of practice. For this reason, I do not undertake the mapping of the analysed publications on this heuristic. Instead, in the next chapter of my thesis, I draw on the authors' own reflections on their research to develop such a mapping. The chapter 6 of my thesis presents the results of such a reflection by synthesising the findings of the literature analysis with the findings from the expert interviews I conducted with the authors of sampled research articles. With that, chapter 5 of my thesis builds empirical and conceptual foundation for further developing methods assemblages enacted in empirical research on datafication in the emerging field of data studies.

6 Methods assemblages of data studies

In this chapter, I present findings from my synthesis of research materials: drawing on the results discussed in chapter 5, in this chapter of my thesis I turn to the expert interviews with datafication scholars. These expert interviews initially draw on the research projects reported in the research articles sampled for my literature analysis, reflect on these studies, and touch upon other projects researchers pursued in the past, in the present to the date of the interview, or planned for the future. Therefore, rather than being an independent study, findings from expert interviews presented here build on and grow from the analysis and findings discussed in chapter 5, complementing these, broadening the perspective, adding to them not only additional detail omitted in published research articles, but also the active reflection of datafication scholars on their research. Together, chapters 5 and 6 comprise the literature synthesis I conducted with my thesis. The following quote is an illustrative example of the way in which expert interviews expand on and deepen the analysis of published articles. Rephrasing an interview question I asked, one of the experts summarised their methodological approach to studying datafication: "How do you determine what you're studying? That's the question, right? Most of the time I have a hunch. And I think this is going to be interesting" (I24, pos. 6). Such 'hunches' can be difficult to account for in academic articles or even difficult to put into words. Reflecting about what 'a hunch' means, how it configures research goals and designs, what expertise and specific research procedures help developing a hunch in a successful academic study were, therefore, in the core of the expert interviews.

Following datafication experts' reflection on their own research projects, I attempted to draw connections between the multiple elements of the methods assemblages as these were enacted in empirical studies on datafication discussed in the interviews. By exploring these connections, I reconstructed how these methods assemblages allowed interviewed experts to develop empirical, re-situated conceptualisations of datafication processes they studied. The aim of my analysis presented in the following is, thus, to understand, what is, according to the interviewed scholars, interesting about datafication processes and how methods assemblages turn 'a hunch' into new, situated concepts about datafication. These assemblages can be placed within the heuristic developed from my literature synthesis and are distinct from one another in relation to the kinds of actors addressed in the studies of datafication processes (individual or collective actors and their positioning on the continuum between 'users' and 'producers'). Furthermore, as my analysis of expert interviews suggests, methods assemblages can be distinguished from one another in relation to the empirical processes that are conceived of as 'datafication' in empirical datafication scholarship: some experts primarily explore data infrastructures, others attend to various data representations in form of data visualisations, metrics, and numbers and how people encounter these; finally, some investigate datafied regimes inscribed and enacted through algorithms and policies regulating their design and/or usage on various levels (e.g. corporate and state).

Empirically, methods assemblages gather together various elements of a research process, including relevant actors (researchers with their professional and personal histories, researched individuals, groups, and/or things such as documents or technologies all located within research situations spatially and temporally (Zakharova, 2021). These elements are assembled together through research practices including procedures and techniques of data collection and analysis such as interviewing, practices of organising research processes, and practices of communicating research

to academic and broader audiences. In sum, methods assemblages can be understood as analytical lens to describe how *ordering and associating of human and non-human actors* is enacted in the research process.

Table 6-1 illustrates some aspects of such orderings specific for each of the three methods assemblages constructed through my analysis: who and what is being included and excluded, othered from the methods assemblage. The question of '*what we are talking about when talking about datafication?'* addresses the core of datafication processes—the means by which multiple realities are being translated into data and back in a given empirical case. Depending on these means of enacting datafication processes, the units of analysis that data scholars explore in their projects vary from encounters with data representations, to dynamic, changing implications of data infrastructures, to datafied regimes inscribed in algorithms and regulating them policy documents. Finally, each of the constructed methods assemblages can be distinguished from the others based on specific methodological characteristics. For example, scholars exploring people's encounters with data representations often create or recreate these encounters within their research projects to learn together with their study participants. Researchers interested in tracing dynamics of data infrastructures follow their units of analysis over long periods of time, exceeding the usual duration of a single research project. Scholars aiming to understand datafied regimes inscribed in algorithms or policies regulating their design and usage follow their interest in practices of data production and processing within which datafication processes and representations of digital data are being negotiated.

The aspects highlighted in table 6-1 refer to the definition of the methods assemblages I provided in chapter 3, according to which methods assemblages are comprised of the researchers and their positionings, the researched, the empirical sites of practices these researchers explore, as well as research processes, practices, and techniques. So, the question of *'what we are talking about when talking about datafication?'* renders visible the means through which datafication processes are enacted at the *empirical sites of practice*. Units of analysis also give insights about the studied empirical sites of practice by highlighting which kinds of actors are being assembled together through the methods assemblages. Finally, the methodological characteristics listed in the table relate to the research procedures and to the theoretical, epistemological assumptions underpinning empirical research on datafication. In sum, my analysis of the interviews indicates that methodologically, current research about datafication in social sciences can be understood with the help of three kinds of methods assemblages: 1) an encounter with data representations, 2) tracing dynamics of data infrastructures, and 3) reconstructing datafied regimes. I distinguish between data representations, data infrastructures, and datafied regimes inscribed in algorithms and policies regulating their design and usage. This distinction is drawing on the analysis of my research material and reflects my sample. In empirical practice all these are often interrelated. For example, digital data generated through data infrastructures might be visualised in diagrams or put into tables (data representations) that other actors subsequently encounter in their everyday or professional lives. The negotiations about how these digital data and their representations are attributed meaning can be a part of a datafied regime and inscribed in certain policies. My goal here, however, is to illuminate how different methods assemblages, addressing different *kinds* of datafication processes, produce different conceptualisations of what datafication is or can be. I provide here an analytical, conceptual way to address datafication processes by distinguishing between these three answers to the question of what we are talking about when talking about datafication.

Table 6-1 Methods assemblages: an overview

Methods assemblages	What are we talking about when talking about datafication?	Units of analysis	Distinctive methodological characteristics
Encounter with data representations	Data representations (e.g. visualisations or rankings, ratings)	Experiences of encountering data representations	Learning *with* study participants by creating encounters
Tracing dynamics of data infrastructures	Data infrastructures	Data movement across infrastructures and their implications	Long-term focus on data infrastructures
Reconstructing datafied regimes	Algorithms and policies regulating their design and usage	Datafied regimes inscribed in algorithms and regulating policies	Interest in negotiations about and practices of data production

An encounter with data representation, as the title suggests, elucidates the roles of data representations such as data visualisations and numerical representations, e.g. in form of numbers, tables, ratings, or rankings, in datafication processes. While data representations as a part of research on datafication can be considered as a research technique in itself used not simply to visualise but also to analyse digital data, this is not how I address data representations here. Rather, I address data visualisations, tables, and ranking as visual or textual representations of digital data that other actors negotiate, encounter, produce and relate to in their everyday or professional practices. I do not focus on the creation of data representations as a technique used by scholars and other actors to enact or understand datafication processes: as my analysis in the following sections of this chapter will show, my research materials do not provide enough reflection on this. Instead, I understand data representations as artifacts created and used at the empirical sites of practice, by means of which datafication processes are being enacted as a part of other empirical phenomena and practices. The methods assemblage of tracing the dynamics of data infrastructures, in turn, attends to the historical, social, political, economic, and cultural aspects of big technological infrastructures used to move data between and across various actors, while the concept of infrastructures here draws on the work by Leigh Star and Ruhleder (1994). Finally, the methods assemblage of reconstructing datafied regimes addresses algorithms as central components of datafication processes. As I show in the following, some scholars draw on other units of analysis, such as policies regulating usage of algorithms in order to reconstruct their datafied regimes. The algorithms and such policies are considered together as researchers studying both aim to understand how certain data representations are negotiated in these. Besides, as findings from my literature analysis in chapter 5 indicate, in my sample, scholars exploring datafied regimes rather rarely study computational rules of algorithmic systems; a more detailed analysis of that is presented in the following. Identifying whether the methods assemblage draws on an investigation of data representations, data infrastructures, and algorithms allows to refer a methodological approach to one of the three methods assemblages in regard to the human and non-human actors which this assemblage draws together. Following the reflections of the datafication experts in my analysis, in this chapter I also draw on the wording the interviewees use for discussing their units of analysis; the terms used for defining the three methods assemblages as well as terms used in table 6-1 present my interpretation of these reflections.

Research design of interpretive and explorative studies that constitute the majority of my sample is guided by the specific research question and attending to specific units of analysis. Hence, the second characteristic of methods assemblages addresses what kinds of aspects of datafication processes were explored empirically. For example, an encounter with data representations takes

place either in the everyday or professional lives of people or as interventions planned and staged by scholars. Researchers, then, can explore how different people experience such encounters with data representations. The methods assemblage of tracing the dynamics of data infrastructures acknowledges the long-term impact of datafication processes on groups of people (e.g. defined by region or affiliation to an organisation) and elucidates historical, economic, cultural, and political implications of datafication processes for different stakeholders. Finally, other scholars, aiming to understand which datafied regimes are inscribed in algorithms and policies regulating their usage, attempt to reconstruct these datafied regimes in their empirical investigations, examining, for example, interfaces or affordances of various information systems.

In the following sections of this chapter, I elaborate in more detail on the three methods assemblages. Each section is dedicated to one of the methods assemblages based on an analysis of expert interviews with scholars who conducted empirical research on datafication processes. In order to reflect on each of the methods assemblages, I begin this chapter with an exploration of epistemologies and politics of data studies as seen by the interviewed experts. The section 6.1 of this chapter, therefore, illuminates experts' philosophical assumptions, subjectivities related to the methods assemblages of data studies, and their takes on research politics of data studies, including issues of interdisciplinarity and publishing. The sections 6.2-6.4 report on the construction and my conceptualisation of each of the three methods assemblages:

- 6.2 – methods assemblage for studying encounters with data representations,
- 6.3 – methods assemblage of tracing dynamics of data infrastructures,
- 6.4 – methods assemblage of reconstructing datafied regimes.

I conclude this chapter with a discussion about what is assembled with methods assemblages and how do they produce different re-situated conceptualisations of datafication processes (section 6.5).

6.1 Researching datafication: politics, emotions, and epistemologies

Before attending to the individual methods assemblages currently used to inquire about datafication processes in social sciences, I turn to the question, why the interviewed scholars study datafication and how, for them, datafication research is different from other academic domains.

6.1.1 Data, research, politics

First, digital data and, respectively, datafication processes are highly political as discussed in chapter 2 (e.g. Prietl, 2019; Ruppert et al., 2017). Several of the interviewed experts underscore the role of data and datafication processes for political decision-making, for example in regard to the availability of resources for development of technological infrastructures, qualification of relevant personnel, and negotiations of data representations (e.g. I13, I17, I19). For many researchers, political actors (e.g. government officials and other public authorities), therefore, are among the stakeholders relevant for their empirical inquiries. Political actors are involved in studies of datafication processes in multiple ways. For example, some of the interviewed scholars study datafication processes enabled by public authorities or the implications of these, for example by attending to policy documents (e.g. I2, I4, I5, I6, I15) or directly exploring political decision-making processes in which data representations are being negotiated (e.g. I5, I29, I30). Second, some of the interviewed scholars entered partnerships with public authorities during their research that allowed them access to the sites of practice (e.g. I13, I30). Such partnerships were often built on requests for collaboration that the scholars received. In both cases, scholars studying datafication empirically depend on access to information, people, and data relevant for datafication processes enabled by various political actors and maintain mutual relations with political stakeholders. For example, some interview partners recount how political situations in the regions of their work such as changing

political parties and agendas, led to delays in their empirical work (e.g. I13, I17). I17 recounts their[19] experiences with policy-makers at an empirical site of practice as power struggles literally unfolded during their observations and empirical data collection phases. These political struggles and mistrust of the stakeholders were also tangible for I17 in the research situations they entered. Alongside these challenges I17 also recounts self-promotion they encountered in their interviews that complicated learning for the goals of research, felt frustrating, and required adjustments in the research design.

Not only are datafication processes political, but, for some actors, empirical findings from datafication scholars provide valuable resources needed to support their political arguments and power positions. Such scholarly findings, therefore, sometimes can be instrumentalised to support certain political agendas, sometimes even regardless of how critical these findings are, as the following example shows. So, I4 recounts several situations in that their works were used to support arguments contrary to their empirical findings: "they did not even make a big effort to rephrase, was taken by an [political actor], copy, pasted to support some statement that that politician was making" (I4, pos. 15). I8 also underscores the politics of datafication processes at the sites of practice and explains how data emerging from various datafication processes are sometimes used by "other people as sort of campaigning tools as well." (pos. 15) To work against such instrumentalization and campaigning, sometimes they develop their research designs tactically: for example, surveys may be important for political work as they produce statistical, quantitative results that "will be more influential" (I8). I18 reflects on certain perspectives on datafication processes by political actors, and respective publicly available reports about these. According to I18, not all public and political stakeholders have required methodological expertise that would allow them to explain and critically reflect on various datafication processes such as use of algorithms for governance or policing. I18 elaborates on the politics of methods required both to understand and to present insights about complex datafication processes, particularly in 'grey' literature, published not directly by academic.

> "[H]ow can we have a conversation with stakeholders, like policy stakeholders about their methods? And because they're not robust a lot of the time. I mean, some of them are, some of them are great. […] But, you know, there are others that […] It's misleading, you know, in a way that I don't think is intentional. And I think that is to do with the sort of lack of sort of methodological expertise and a nuance" (I18, pos. 25).

I18, therefore, identifies methods and methodological expertise for policy stakeholders etc. as an important discussion topic in the partnerships between academia and practitioners.

Another expert, I6 gives an example of challenges in navigating politics of datafication processes. They explain their caution in analysing results of interviews with political actors, grounded in a lack of openness in communication between them and the researchers. Challenges in communication between political actors and datafication scholars can have multiple reasons. For example, I17, speaking about European public actors processing personal data mentions that "they do not provide it [data] voluntarily" (Pos. 15) touches upon challenges in acquiring access to digital data for research. I19 gauges the possibilities to access relevant stakeholders, information systems, while I15 turns to the analysis of policy documents to "substitute the difficulties of field access" (pos. 9). Finally, I6 addresses challenges in communicating with political actors (also applicable to other practitioners) due to the different languages they and researchers sometimes speak:

> "And then the other issue is that, because a lot of these technologies are so new, and the language around them is so new, […] there is a communication challenge. Because sometimes, you know, what we would call predictive analytics at the time" (pos. 9).

[19] For the reasons of anonymisation, in this and following chapters all interviewed experts are addressed with the pronoun 'they/their'.

Talking about partnerships with external communities and organisations, I26 reflects on critical research about datafication processes and findings that complicate rather than simplify these processes:

> "it's not necessarily something that certain people want to hear. Like for example, if you're thinking about setting up a partnership with an external organisation, do they want to hear how messy and confusing data is?"(I26, pos. 27)

Scholars working in other countries than their institutions such as I4, run the risk of being accused of having an outsider perspective that is not relevant to the local communities. Recognising values and needs of local communities, indeed, is important for navigating the politics of datafication processes, especially in regions that suffered under colonial oppression. Working together and helping "NGO friends" (I4, pos. 23), is a way of engaging with local communities for some of the interviewed experts.

Engaging (local) communities at the sites of empirical practice and attending to their needs, values, and imaginaries in regard to the datafication processes is not only part of research in various non-western countries. Many of the interviewed experts followed a participatory approach in at least one of their projects discussed in the interviews. That means, communities of stakeholders at the empirical sites of practice were involved in the research processes not solely as informants, but also as decision-makers and beneficiaries of the research processes. For example, the goal of the I11's project was to develop software for a group of practitioners whose requirements and imaginaries of the software were in the core of the research and development. In another project, I20 and colleagues developed software that should be used to "create opportunities for non-experts to learn […] about [technologies]" (pos. 15). I23 and their colleagues were also concerned with providing their study participants with information available to the research team about the software they explored. For some of the interview partners, therefore, enabling their study participants, particularly laypeople, in the face of various datafication processes was one of the central research goals.

6.1.2 Persona of a researcher, emotions, care, and critique

For some of the interviewed experts, their personal, professional backgrounds and histories allowed to develop approaches in which they shared their power positions as scholars with the study participants, as it is central for participatory approaches (e.g. Vines et al., 2013). So, I10 tells about their previous background as a practitioner in the same field they currently study. According to I10, it helped them in their research to reflect on their own standpoints and views on datafication processes and resulted in an action-research driven study design that allowed to act according to these personal assumptions. They described their "love-hate" relationship to data.

> "The most interesting thing about it was that I discovered that at times, I loved it. So, I wrote in my field notes, I love data. I also hated it and felt angry with it and thought this is ruining [the empirical site of practice]" (I10, pos. 16).

Originally being critical towards datafication processes, their views changed over the course of their research project, as they recognised data as an essential part of empirical practices they explored.

A few of other interviewed datafication scholars share such conflicting views on datafication processes. For example, I2 "tend[s] to be normative and judge" (pos. 19) digitalisation and datafication processes as these reify multiple inequalities. They also notice how their latest academic projects made them reflect on their personal, emotional experiences of datafication processes.

> "And then, I was also reflecting on my personal situation as well. Because sometimes we don't realise how many things, we take for granted and how much like data is kind of fostering some competition. And this kind of rhythm of life that we live nowadays, it's kind of disseminating these ideas like of competition through data, and through the indicators, or rankings, or everything. So, there is the idea of being more efficient, having evidence for

> making decisions, which is good. But then behind that, there is this idea of competing, of trying to be your best, or trying to reach the ends no matter what the means are. And these ideas are somehow like a bit violent to every day's life. And so, this makes people unhappy, stressed, and sometimes frustrated. And this was kind of the way in which, like, I was surprised by my research to realise how much does this really present in the daily lives of like myself, but also the people I interviewed" (I2, pos. 23).

I22, shares I10's love-hate relationship with digital data and I2's reflections on their research:

> "I mean, it's such a weird world, isn't it? This metrical world, everything being ruled by metrics. Yeah, I guess as a qualitative researcher, I fundamentally have a high value for asking those why questions, but certainly through working on datafication, I've increasingly come to see the value of having some numbers, but it's hard work getting them in terms of research design" (I22, pos. 23).

Similarly, I7 started their research projects assuming a critical position towards datafication and quantification. That position, however, slightly changed in the course of the study after facing some positive examples, although I7's overall critical perspective was confirmed in their findings.

I27 shares another story of conflict between their personal views on datafication processes and their work. Prior to their career as a researcher, I27 worked in the field they currently study, which led to the development of their research interests. Reflecting on their previous practical experiences and on

> "how far away the data that we were collecting, and the picture of human beings in [field of study] that those data could tell us was from […] my spirituality. And the answer was a lot" (I27).

They address emotional discomfort and disappointment that served as a springboard for an academic inquiry tending to the experiences of their former colleagues (I27). While I27 foregrounds their spirituality, I18 notices how personal standpoints towards datafication and, respectively, the concepts about datafication produced within current critical datafication scholarship are also tightly intertwined with political views of critical scholars. I18 also elaborates on the role of critique and the interplay between civic and academic critical positions.

> "You know, it's quite a tricky thing to talk about, because it's quite a politicised field of study. You know, there is quite a lot of sort of left leaning critical thinkers in the field, I'm left leaning too. And you could arguably say that you find what you are looking for, you know, you find that data-driven systems discriminate against certain groups of people, or you find that managers are using data tracking technologies in surveillance ways, or you find that people don't like things the way they are and would like them to be better. You know, I'm not sure how willing people are to hear that about their own research. But I think it's a phenomenon there is not enough reflection on" (I18, pos. 11).

For other interviewed experts, such as I8, personal experiences such as being a parent also shaped their research practices. Being able to observe datafication processes in practices and experiences of people directly impacted by these processes not only as a scholar but also as a stakeholder, and a parent: "it gives me a more rounded view, I suppose, of the process" (I8, pos. 17).

This shows that the personal standpoint of researchers does not only mean offering critique of certain concepts, but also the passion and care of researchers towards their work, other people, and the world, through their work. So, for example, discussing their research design, I1 mentions how they like conducting qualitative research, as it allows them to become involved in interesting things (I1, pos. 7). I5 sees their role as a datafication scholar not in labelling certain processes as good or bad, but in understanding their variations and "still trying to give hope" (pos. 55) to the communities they work with. I30 calls for approaching datafication processes "with an open mind and find some nuances" (pos. 15). Despite such hopeful and caring standpoints towards datafication processes, most of the interviewed experts would agree on what I3 summarises in the following way:

> "So, maybe there is still kind of constant things that seem to be-, seem to annoy me and my, some of my colleagues. Also, to know, and still seem to annoy something-, one of the things is that this-, because these are people who are kind of proposing solutions to problems that they see in the society at the moment. But somehow, those problems are really heavily fixed on the data and datafication kind of plays out at the moment now. So, sadly. So, it is, really difficult to think about anything else for these people than to just, you know, figure out how people could use their data, own data for their own benefit" (I3, pos. 19).

Other experts express a range of negative emotions they experienced during their studies of datafication processes. I13 reflected on frustration that accompanied some phases of their research projects when no empirical work could be conducted (and no results produced) due to the delays at the empirical site of practice. While they waited until certain data infrastructures could be further developed "there was very little that we could write about" (I13, pos. 15). For I21, who studies social media, doing extensive research on a particular platform makes using the same platform personally tiresome and affects their feelings about particular apps. These personal experiences of datafication research are intertwined with personal stories and histories of the interviewed scholars, with their onto-epistemological, and 'spiritual' assumptions, and their positions in the academic and other communities. My findings indicate, in the spirit of feminist technoscience (e.g. Haraway, 1988; Suchman, 2002) how articulating these assumptions alongside with research findings produces more situated, contextualised accounts which render visible researchers' care for their objects of study.

While some of the negative emotions expressed in the above quotes are quite specific to studies of datafication, digitalisation, or mediatisation processes, the interviewed experts also shared their emotional experiences common for various academic domains. For example, I24 sometimes experiences writing with others as frustrating as they feel like they do most of the work. I12, in turn, was frustrated about lacking support and challenging communication in an interdisciplinary team. Particularly these latter examples refer to the significant role of scholarly communities for the individual scholars, and ultimately, for the production of empirical findings and conceptual understandings of datafication processes. It is specifically relevant for junior scholars like I12 at the time they mention in their interview. For I10, being a part of an academic community shaped their understandings of datafication that extended on the work of established scholars in their field.

> "And my work really builds on their work, you know, that they are the giant shoulders that I'm standing on, you know, because they've done such a lot with it. I think what I was doing was theorising it a bit more. They have documented it, but I'm interested in theory" (I10, pos. 10).

For senior scholars, their academic communities provided the necessary springboards for developing new ideas (e.g. based on gaps in literature) and the necessary support for political work that many critical scholars perform in their respective regions of study (advising, critiquing, and developing ideas) especially concerning the increase of datafication processes. While engaging with the works within the academic community, the ontological differences need to be acknowledged. For instance, I15, an associate professor, is

> "[t]rying always to understand what earlier research has said about the topic that interests me, but at the same time, I'm aware that that research might be coming from a very different [field]" (pos. 7).

Similarly, I18 (a full professor) elaborate on how they "decide what questions are important, based on current debates, and gaps in those debate" (pos. 3). Similarly, I25, also a university professor, while planning a research project, is also focusing on societal, academic, and political needs. Other experts view their empirical research as practical contributions to their respective scholarly communities. For example, they can provide other scholars with open-source resources (e.g. crawlers, datasets) for further studies (e.g. I6, I14). Overall, communities allow datafication scholars to collaborate with other researchers with shared understandings of datafication processes,

developed through literatures. However, as I24 notices, it can also limit their perspectives on datafication processes. For example, I24 laments, in their domain-specific academic community there is yet not enough conceptualisations of datafication that would acknowledge the particularities of the studied empirical sites of practice. Rather, the community borrows concepts from other broader research domains, which do not necessarily translate well. The lack of domain-specific conceptualisations of datafication, on the one hand, can be rapidly developed through the studies of multiple empirical examples not yet discussed in academic publications, but on the other hand, requires theory-building from datafication scholars in every publication, including empirical datafication research. Further challenge that I24 sees for their research community is its small size. They notice how academic trends are developing fast and influence how findings are presented, meaning that "almost everyone does the same" (I24, pos. 14). I25 develops this point further and explains on an example from their teaching how in such smaller, emerging scholarly communities research findings are written and presented with many assumptions. These assumptions are well-known within, but not necessarily outside the scholarly communities, as I25 explains in an example from their teaching practice.

> "The students didn't understand it was a critical text. They thought it was a text explaining how [this software] works and I thought, what? It is a throughout critical text. And what I find particularly interesting is that we… […] We are writing with so many assumptions that are not clear for students. So, when Ben Williamson is writing about big tech or Silicon Valley, for me it is clearly a negative connotation. But they [students] do not understand the negative connotation" (pos. 17).

The example provided by I25 also illustrates another theme surfacing in most of the interviews I conducted. This theme addresses the role of critique in datafication scholarship and gives some answers to the question of why study datafication processes in the first place? While most of the interviewed experts describe their research as critical, a few of them underscore the importance of generative critique in datafication scholarship. For example, I20 recognises the importance of critique of datafication processes while at the same time elaborating on how "dismissing that without understanding how that technology not only functions but is already being utilized" (pos. 13) is not helpful. They assert that critique should be grounded in the understanding of technology to live up to the pace of technological developments:

> "if you're talking about datafication I'm sorry. But like just a kind of social or cultural critique just, it's not enough, it's going to exist on the surface and it's going to miss- even let's say, just understanding what's different or meaningful about a shift from a kind of symbolic to a connectionist AI model" (I20, pos. 15).

I3 and I25 also articulate the need for more generative critique in the datafication scholarship:

> "So, what should we do differently? And so, kind of figuring that out, has been really important to me. And it goes to this expertise in the sense that, it is really easy for us who read these critical [social] studies literature to figure out the problems. It is really easy to point out the kind of the-, do criticism, I would say. So, that is kind of-, I am not saying it is cheap. But it is easy sometimes because in many cases, kind of you already know what you are going to find […]. So, you are going to find things that you often find with these kind of data related initiatives like you will find capitalism, you will find exploitation, you will find kind of surveillance, you will find biopolitics, you will find the entrepreneurial self and so on. So, there is this kind of concepts that you know, we can apply the things that we do research on and then you kind of already a bit know what is going to happen and that is-, I am not saying that doing that stuff is not important, but I-, it is would be a specific kind of expertise to do something else. So, be kind of still sort of within this critical literature, but still not just do critique, but be productive." (I3, pos. 25).

> "To find a way… See, I cannot even describe it for you. Positive examples about datafication. […] Find positive examples or how people repurpose data for themselves and do something with these. Without talking about best practices. Or naïve endorsement. I think generative critique like, like I mentioned in cultural studies, is that through highlighting what the like they criticise the whole society. […] But it is difficult in [I25's] research domain, because it can be easily misinterpreted as a good practice, best practice, next practice. We need to do it like that, it's cool like that. And for me that is a big challenge. How to write about it? About data, datafication, data practices, without such reduction." (I25, pos. 31).

Similarly, I5 states that critical scholarship is not 'just' about offering critique, but also about trying to understand and capture what is going on at the empirical sites of practice. In this spirit, I30 aims to extend current research about datafication by looking at practices to produce a more heterogenous picture of datafication processes. They also view datafication as strongly politicised, not at least in academic discourses. According to I30, these political discourses are either for or against datafication, though the middle path exploring the diversities, complexities, and multiplicities of datafication processes would pay respect to the different articulated positions and the heterogenous social practices through which datafication processes are enacted.

6.1.3 Epistemologies and institutional frameworks

These critical-generative perspectives on studies of datafication processes are interconnected with the theoretical and onto-epistemological concepts applied in datafication scholarship. The interviewed experts working across different domains of social sciences, attended to various concepts about datafication described in the previous chapters of my thesis. Some of the projects discussed in the interviews began at the time when "Mayer Schönberger and Cukier's [(2013)] book had come out. And right around when Jose Van Dijck [(2014)] had written her seminal piece on datafication" (I14, pos. 5) and were widely discussed in the field. These discussions and early conceptualisations of datafication processes facilitated the development of empirical projects mentioned in the expert interviews. Further conceptual frameworks such as surveillance were used to support the empirical and theoretical findings from the discussed projects (e.g. I15, I29). I29 gives an example of readings relevant for their current projects: "I've just finished a few months ago reading '[The age of] surveillance capitalism' [(Zuboff, 2019)], and it really articulates things that I was thinking before" (I29, pos. 15). Such theoretical frameworks set the boundaries for what to include into or exclude from the concept of datafication. However, even with a number of theoretical and onto-epistemological perspectives on datafication processes "behind" them (I8, pos. 23), datafication scholars still face various conceptual challenges. I24 describes the concept and term 'datafication' as "slippery" (pos. 4). For them the term datafication is "a shorthand for so many different things at once" (I24, pos. 16) and cannot be distinguished sufficiently from other terms such as 'digitalisation' or, sometimes, even 'technology'. Similarly, I26 elaborates on datafication as a concept standing for many things difficult to pin down even in one societal domain, not at least as there are also quite different imaginaries of datafication both among researchers and practitioners. I14 is specifically concerned with the empirical and practical questions arising from the concept of datafication such as

> "how is this sociality expressed through the technical? But then with the kind of added question of like, if sociality is being expressed through that, then what are those technical objects that is allowing for that expression in the first place?" (I14, pos. 5).

The quotes from the interview illustrate that for scholars studying datafication processes, the term and concept of 'datafication' also means different things: societal processes and imaginaries of various groups of people including academic communities themselves. While I14 addresses empirical challenges of pinning down what is datafication, and I24 or I26 point out to the conceptual challenges, I15 reflects on this double meaning of the term 'datafication' as an empirical process and a concept, theoretical framework used to study these same empirical processes.

> "Well, one thing that I'm struggling with is I think kind of datafication is both a concept that we work with, but it's also kind of a phenomenon that we're studying. And I think that makes it really difficult. How do you -, because I think for me it's an approach. Or like I would say that, datafication is the kind of lens that you apply to a phenomenon that somehow is -, you see it as related to datafication, but maybe it's not correct to just say, okay, that is datafication, that is not datafication. So, but for me, that kind of always keeping in mind that there is this confluence between the conceptual and the empirical. I think that's something that troubles me. And I don't know how to sort of resolve that. But that's one of the, maybe tensions that I am feeling. How do we still kind of use it conceptually as opposed to -. Or kind of conceptually empirically, but also understand that it's a concept, that it's a kind of lens that we are applying." (I15, Pos. 17).

I17 also grappled with the same question about the concept of 'datafication' that led them to thinking about digital data as "neither persons nor objects, but something in-between" (pos. 11). Avoiding the binary of subject-object understandings of digital data, I17 views them as an ontologically other manifestation of reality.

In hindsight of these challenges, the choice of further, domain-specific theoretical frameworks applied in each individual study becomes a significant part of the study's research design. Several of the interviewed experts articulated their interest in concepts stemming broadly from STS (e.g. I15, I21, I30), while others offered criticism of the same. I15 summarises their reflection on the use of STS in datafication research:

> "But maybe then, that also makes it a little bit difficult sometimes because, then you're really kind of working in this nice bubble that provides you all of the resources you need" (pos. 7).

Overall, STS-inflected datafication scholarship remains one of many theoretical, onto-epistemological traditions used to understand datafication. I19, for example, view their work as grounded in interpretive (German "*verstehende*") sociology and follow explorative empirical approaches to map various "puzzle tiles" (pos. 3) of the complex datafication processes performed in society. Within that explorative approach, I19 strongly focuses on individual cases, therefore making ethical questions of preserving anonymity of the study participants ever so important as they place a lot of trust in the expert. I20 prefers applying cultural theories (e.g. as those developed by Innis and Foucault) rather than STS, as these allow I20 to better understand the politics of datafication processes. Other experts such as I8 and I31 also extend their research on Foucault's theoretical approaches. The different approaches interviewed datafication experts apply to understand their empirical findings, however, do not complicate or trouble a conceptualisation of datafication. Rather, applying theoretical frameworks that speak to the relevant research domains and empirical sites of practice allows to produce a more situated perspective on datafication processes than a universal theoretical framework (if such would exist) could ever do. As I24 argues, for future research on datafication, more theory-building should be done.

According to the interviews, for theory-building in empirical datafication research certain methodological choices need to be made about the analytical frameworks and lenses through which to address datafication processes. The interviewed experts make such choices differently. First, some researchers apply theoretical frameworks according to their pervious expertise and research interests. For example, I1 applied frameworks from gender studies while I10 turned to postcolonial approaches. In the beginning of their empirical project, I10, who also has a background as practitioner in their current field of study, started reflecting about their position of power as a researcher, which led them to choosing a postcolonial and participatory research approach that balanced differences in power of empirical stakeholders, study participants, and the expert. I13 and their colleagues used a philosophical framework they considered inspiring and got feedback on their use of the theoretical approach from colleagues in other disciplinary domains who work with the same framework more often. I13 also explicitly acknowledges the performativity of epistemological and methodological frameworks chosen for empirical research and offers critique of "ethnographic

positivism that basically you go in there and you capture everything, and you detail everything" (pos. 27). Second, other experts' theoretical choices were partially influenced by the editors, inviting scholars to contribute to anthologies and edited special issues of academic journals (e.g. I1, I30). Third, in consortium research projects choices regarding theoretical, onto-epistemological, and methodological approaches to research are often made prior to the study through grant proposals. To adhere to the goals described there, scholars later working on such projects have freedom in choosing different field theories, though are constrained and guided by the overall project frameworks.

Consortium research projects often define the overall research framework (e.g. I2, I12, I18, I20, I26, I28), although new topics can emerge as findings. For example, I2 tells how in their consortium, different research groups conducted their own studies, but also did joint analyses to relate the results to each other and bring them in dialogue within the project. In this consortium, senior scholars who applied for funding pre-defined the initial study design, while other scholars involved in the project could make their own conceptual choices (I2). Another expert, addressing the advantages and challenges of working in a consortium project, notices that "one of the things that informs how we do things, is the teams that we are in" (I18, pos. 7). I28 makes a more general note stating that "science is a team sport" (pos. 13), therefore different people are required as there are many tasks to accomplish in one project. For managing the heterogeneity of profiles, research interests, and approaches, common research goals and questions are central. While negotiations and talking to each other help translate between different theoretical and disciplinary approaches, working in a team of people with diverging research interests does not necessarily yield into "completely perfect" (I18, pos. 9) research designs according to I18 (also I20, I26). For example, I26 articulates their experiences of working in collaborations in which strong research profiles of team members, being central to the development of studies, also made it difficult to contribute alternative methods and produced interesting, though not quite sufficient for everyone involved, results.

As several interviewed experts underscore, despite the challenges of working in bigger or interdisciplinary[20] teams, it contributes greatly to understanding datafication as a multiplicity of heterogenous processes that require analysis from a variety of perspectives: "different kinds of eyes looking at the forms of evidence" (I29 pos. 19, also I19, I25). Making a similar point, another expert (I20) addresses datafication as a societal phenomenon not bound by any discipline, which, therefore, needs to be studied from an interdisciplinary perspective. Some experts, such as I6, underscore both the value of an interdisciplinary perspective and that of seeing implications of datafication processes from a disciplinary perspective.

> "Because I think there's real value in an interdisciplinary perspective. And there is also real value and interrogating, you know, the implications of datafication from a wide range of disciplinary perspective" (I6, pos. 23).

According to the experts, with the help of such interdisciplinary approaches, the complexities of datafication processes and the interrelatedness between various aspects of these processes can be rendered visible. Another advantage of bigger, interdisciplinary consortium projects for studying datafication is their funding. I26 summarises it as follows:

> "But in reality, I think I've found over the last few years that unless you've got a lot of funding for massive project, which is increasingly more challenging to combine, then really, you're very constrained in terms of what you can do" (I26, Pos. 5).

The same expert also mentions that, especially for datafication research, topic choices are also highly political, and funders also pursue their own agendas, resulting in granting or denying funding for certain projects.

---

[20] Here and in the following, I do not make distinctions between inter-, trans-, or multi-disciplinary approaches but rather use the term "interdisciplinary" for all the above. This decision is grounded in a variety of uses of these terms in the expert interviews.

> "I mean, I guess if you were centering a funding application on that issue [(datafication)], maybe it'd be something that they'd value. It probably depends which funder you're talking about. They just might, some reviewers that wouldn't see that, wouldn't appreciate it might just think that you're incompetent like my technical reviewers or something, who knows. So yeah, I think you probably do need to be a bit careful about how far you go with discussing that in different contexts." (I26, Pos. 27).

I20 recounts the procedures of funding applications for different research domains—social sciences and engineering/computer sciences, concluding that the choice of funder shapes the research project. The arguments about the role of funding put forward by I20 and I26 can also be related to the arguments made by other scholars (e.g. I8, I18) who highlighted political aspects of digital data and of research about these (see section 6.1.1). These latter experts mentioned how in developing research designs and questions, they follow both the current academic and political, societal discourses, while the choice of research techniques sometimes is driven also by the need to give a certain argument more political 'weight' (I8). For I19, developing multiple funded research projects iteratively allows to gather the "puzzle tiles" (pos. 3) of one bigger, complex datafication process.

In general, most interviewed experts understood their work as interdisciplinary, while some specified that they work with other social scientists (I18, I19, I29) and others with computer science and other natural and engineering sciences colleagues (e.g. I3, I16, I20). While describing their interdisciplinary research projects, I29 noticed how they work in "reasonably interdisciplinary" (pos. 19) teams of scholars from various domains of social sciences:

> "It's fascinating that you asked me that because I fancy myself as an interdisciplinary person and I love working in interdisciplinary teams. But thinking about virtually all the projects I've just told you about, it was basically with people from a similar disciplinary area" (ibid.).

In turn, I25 defines "real" (pos. 27) interdisciplinary collaborations as those in which everyone involved knows beforehand that applied concepts will differ; to achieve some common understandings and learn together, they engage in collaborative reading and writing. I21, in turn, mentions that working in a team, particularly when studying personalised algorithms, provides a way to reach certain "consistency of experience" (pos. 13). Both in the cases when interdisciplinary teams are working together or a team from one discipline is working in an interdisciplinary field, a particular attention to the conceptual choices as well as historical and disciplinary developments of the theoretical approaches taken could be helpful. First, for making authors' assumptions explicit to the interdisciplinary audience. Second, for developing shared understandings and concepts of datafication processes and producing a more partial view on datafication processes within the project teams. Particularly in the teams of social scientists from different domains of social sciences, the challenges of finding common understandings and concepts, although present, were not perceived as grave. At the same time, experts from social sciences working, for example, with computer science scholars, mentioned the difficulties in finding common analytical grounds, the need for translation between the disciplines, and "humility" (pos. 9) required to learn from each other (I20, also I16).

> "So, I am not- the one thing that I've learned is not to expect my colleagues in computer science to see or understand the world exactly as I do or to even have the faintest understanding of the kind of- the social or cultural theory that I might use. In the same way that they would not expect me to be able to give as detailed a technical description of a convolutional neural network as they could give to me. I go to them to learn from them with a basic degree of humility, understanding that the entire purpose of my collaborating with them is to learn things that I do not know right now." (I20, Pos. 9).

I23 notices that particularly for studying datafication processes, sometimes data scientists are tempted to leave expertise from other disciplinary fields without attention, which however, is also crucial according to I23.

> "So if you, …, if you succeed in talking with other experts, then you really have great research projects and great results, which are also useful for society, not only for scholarship. But it's not always easy, because of course- I mean, on my computer, I cannot even run a big dataset. I mean, it cannot manage it. So of course the temptation by data scientists to say, "Okay. I don't need-" Not myself, but your expertise might be there, but still whenever you get a successful research project on these fields, you need to get very different experiences" (I23, pos. 23).

Both I20 and I23 conclude that successful interdisciplinary work is about mutual learning: "stop expecting them to be what you are and start to learn from what they are" (I20, pos. 9).

Other interviewed experts also enact such view on collaborative work in their datafication scholarship. For example, the expert in the interview I25 reports how they could learn more about data analytics and data science methods in interdisciplinary consortium projects. The expert from the interview I16 describes the role of a developer colleague in their research team of primarily social scientists and trust the latter placed in the developer and their expertise to figure out approaches to multiple technical questions from the research project. The social scientists from this team, on the other hand, did not require to develop a detailed understanding of computational processes, but provided questions and situated research in societal contexts according to their expertise in the datafication processes explored in the reported project.

> "But again, I think they bring their own cases, for example, or their own interests, as well. I mean, I will say particular with my computer science postdoc, it's been a conversation about, again, because it's something that's not familiar to me. I don't know exactly what the possibilities are. And what I've been interested in for my project has been to try and critically interrogate existing systems from a technical perspective. And, you know, if it's up to him to sort of figure out like, how can we do that, in a way that actually works, because that's not a straightforward thing to do." (I16, Pos. 23).

Another example on interdisciplinary collaborations provides I12. For them, working in an interdisciplinary team that predominantly consists of scholars from other academic domains than one's own sometimes felt intimidating, particularly as a junior scholar. Another perspective provides I29, addressing the challenges of building up a team with a required variety of expertise. To balance that, collaborating with people who already have experiences and expertise with datafication research can be helpful, as I32 reports.

Overall, while the disciplinary perspectives differ, most important for collaborations, according to the interviewed experts, are shared onto-epistemological assumptions about datafication processes (I19, also I4, I8, I25, I26). Similar experiences in their collaborative work discuss also I14, I23, I24 who, therefore, prefer working in a community of friends, like-minded scholars. Even when projects are conducted as collaborations between scholars from different disciplines, it is important to involve other relevant stakeholders (I26). In addition to translation required between researchers, even if they share fundamental onto-epistemological assumptions, translation work is also required when collaborating with practitioners. As an expert in the interview I6 explains, often scholars and practitioners (e.g. public servants) and also practitioners among each other (e.g. different governmental departments) use varying terms to describe similar datafication processes. Without knowing and using the 'right' terms, communication is difficult and acquired information fragmented.

6.1.4 Expertise, research procedures, and ethics

For datafication scholarship, alongside with theoretical choices setting boundaries for the definition of empirical phenomena under study, methodological choices are performative to the understandings about and concepts of datafication produced in empirical studies. According to I26, who understand themselves as critical data scholar, critical data researchers have expertise for understanding and empirically identifying datafication processes. Even among datafication scholars,

some methods (e.g. computational methods) require more in-depth engagement. Experts applying computational methods such as I16 lament how these are "often, I think, not thought about that these kinds of big data methods are actually theory agents" (pos. 27). Similar arguments have extensively discussed in the chapter 2 of my thesis (see also Lindgren, 2020; Rieder & Röhle, 2017). Based on the expert interviews, several strategies for making methodological choices can be identified. First, some experts (e.g. I6, I8, I11, I15, I22) apply certain techniques of data collection and analysis based on their vast experiences in using these techniques and on the centrality of these techniques for certain academic communities. For example, I8 explains their research designs as follows: "I do datafication with sort of traditional sociological research tools of interviewing, and observing, and surveys, I suppose, but they're very kind of qualitative surveys really" (pos. 23), because

> "that's what I know how to do. And that's what people like me do, you know, in my field. So, there's that. The other one, I suppose more seriously, really is that, in all of them, it was about understanding in depth, people's lived experiences of datafication or the policy that we were looking at. And that kind of depth of data and understanding, I think can only be got out through an interview and ideally one where they're in the work setting, so they can bring in the things that they need, they use, and they can go and get stuff, and they can talk to you about it, and also, that you can get a sense of the setting." (pos. 29).

Grounding methodological choices in the scholarly traditions of relevant academic communities, however, seems to be in contradiction to some other interviews in that the experts assert that they are guided primarily by the research questions and, building on these questions, choose techniques to answer these (e.g. I10, I13, I18, I20, I24, I25, I28). For these experts, techniques of data retraction and analysis never define the development of research projects. Rather, these experts articulate their openness towards experimenting with various methodological approaches that lead to answering their research questions. Further, for some other experts such as I26, applying different methods depends on the collaborations and joint research interests developed from these. Finally, I2 and I26 also underscore the importance of applying both qualitative, quantitative (and computational) methods for datafication research "on our terms as the critical data studies people" (I26, pos. 15) as such combinations allow a better technological understanding of datafication processes. At the same time, I26 warns against putting too much hope in the computational solutions.

> "But I think sometimes computational methods can be good for really interrogating something and understanding that more of a technical level, I suppose, what is going on, what the issue is, and maybe being able to think of ways to adapt the system in order to overcome those issues. But I think we should put too much-. We shouldn't have too much hope for computational solutions, but I think it is useful to not dismiss them in some cases I guess is where I'd stand on that. I think the more that we can bring in the more that we can use to explore these issues, I think is good." (I26, Pos. 15).

Articulating their personal standpoint on methodology, I13 critiques an understanding of methodology as data collection and analysis. Rather, they see research methodology as building relationships with people. I13 explains: "I think, what we do as researchers is, we tell stories, and we build narratives" (pos. 19) and later continues, "I think, you need to really immerse yourself in the context and in the stories that you've collected through the context and in the data" (pos. 21). For I13, science is a creative process in that talking to others (in a 'hermeneutic tradition') is a crucial practice and skill. Similarly, I25 also views research methodology as telling stories that give answers to the research questions. I25 explains their position with the background in cultural anthropology. For I8, in turn, "doing research means going out to places" (pos. 37). For I15, research methodology is an "overall approach in research" (pos. 5); it includes theoretical literature and ontological, philosophical assumptions, according to which scholars 'perform' their research and brings together different aspects of the research practice. (This understanding resembles the

perspective on methods performativity taken in my thesis and is grounded in the expert's current interest in the STS and the debates about the social lives of methods.) Making a similar argument, I30 dismantles a narrative about research processes as linear, underscoring the role of "timing and luck and negotiations" (pos. 3) in developing a successful ethnographic study of datafication processes. Specifically, for some research situations luck lies in the ability to gain access to research study participants and being able to enter certain research situations.

> "So just so you know, gaining access to that was very difficult to come as an ethnographer into a ministry and even once I gained access, I couldn't start doing anything because there was a change of minister so everything, that sort of freezes all actions in a ministry because they wait for the new minister's guidelines and attitudes before they can go on with certain initiatives. So, I had to wait maybe four months from the agreements until I could start this ethnography. So, this is just to say that timing was very important." (I30, Pos. 3).

Similarly, I13 reflects on timing as a challenge for empirical research:

> "I remember going in the early stages and they'd say: "Oh, when you come back in six months' time, everything will be transformed and we'd be just using this and would be doing this", and then we come back in six months' time, and they were stuck at the same thing" (I13, Pos. 17).

Finally, I18 highlights the necessity of methodological reflection in datafication scholarship, as they "don't feel like there any particular method sort of lend themselves, more or less to the topic" (pos. 11). For I18, such reflection on application of any chosen method can be reached by acknowledging the performativity, partiality, and situatedness in academic publications. Critical reflection of one's own research practices, for I18, is one part of expertise that datafication scholars require. In the interviews, experts have pointed out skills and expertise necessary for studying datafication. First, understanding of technological processes enabling datafication and computer science or computational research methods were mentioned by several experts (e.g. I11, I20, I28). I20 notices how "for a lot of people coming from social and cultural theory there is a never-ending suspicion of technology" (pos. 13). They argue that for datafication research to be successful, there should be willingness to engage with the technologies enabling datafication beyond the view on technologies as an incremental problem. I3 makes a somewhat similar argument elaborating on the challenges of conducting critical scholarship and the expertise that is required to provide generative critique. I25 follows an approach of "slow research" (pos. 13) and underscores the value and expertise of attending to individual examples of datafication processes that might have significant implications.

> "There are other people who are working fast. And look at fast policy. And it is really important. I was once criticised for taking a frog perspective. But it's cool, the frog perspective. To slow down and to take only a small piece. And look, what does it have, this small piece, seemingly banal, subtle mechanisms" (ibid.).

At the same time, for example I16 addresses high attention to computational expertise as "fetish" that sometimes can also undermine critical research. Rather, being able to combine multiple kinds of expertise are relevant for datafication scholarship. In this spirit, I30 also elaborate on the role of computational, statistical expertise vis a vis researchers' abilities to ask 'naïve' questions. Similarly, I21 report about teaching their students and junior colleagues to "understand data beyond the hype" (pos. 27), also including narratives created by big tech companies. Overall, the quotes bring into the fore the challenge of datafication as research episteme and an empirical phenomenon, at the same time highlighting how usage of digital data for research on datafication processes brings scholars into a position to reflect their own methodological choices and practices.

The importance of heterogenous expertise underscore a few other experts. For example, according to I19, heterogeneity is one of the central kinds of expertise for studying datafication processes, for the expert, it is not helpful to study datafication only from one perspective. For I11, the core expertise of datafication scholars describes their ability to cover various sub-aspects of

datafication processes, including both domain-specific and datafication-specific concepts and implications. Similarly, for I18, knowledge of an empirical field of study also can be understood as valuable expertise. I26 brings all the above perspectives on the expertise of a datafication scholar together. For I26, various disciplinary perspectives are equally important. For example, "some understanding of the information sciences is quite valuable" (I26, pos. 15) to avoid exaggerating the power of the technologies. However, they later notice, these discourses are currently "getting better" (ibid.). They continue,

> "I think studies of datafication can come in all different shapes and sizes, methodologically disciplinary focus, and I think that's really valuable in what creates a richness of the field with everybody looking at things from different perspectives" (I26, pos. 15).

In turn, I24 argues that for datafication scholarship as a relatively new academic field, good understanding of theory is required for theory-building. Besides theoretical expertise, communication, and "people skills" are also relevant as for I24, attending to datafication processes also means attending to their implications for various stakeholders. In addition, communication skills, according to I24, are relevant for building scholarly communities in this fast-developing research domain. Another kind of expertise and skill mentioned by several interviewed experts concern academic writing. For example, for some of the interviewed scholars, writing practices are seen not only as required to produce academic results, but as a way of creative engagement with an object of study; sometimes writing is also combined with walking (I10). Within consortium projects, writing is seen as a valuable form of communication with colleagues through which translations between disciplines and theoretical approaches can be performed and accomplished. However, writing up research findings for academic journals also requires finding "the right" journal that "understands" own ideas and ways of doing things (I10, pos. 22, also I25, I28). In contrast to journal articles, books (e.g. as reports at the end of a consortium research project or a PhD thesis) provide space for a more detailed methodological and conceptual accounts, reflecting the journeys research projects undertook (I2, I3). Several interviewed experts (e.g. I2, I27) expressed their general frustration with journal publishing and academic systems, noticing the lack of time for reading longer pieces and the following demand of shorter texts, poor descriptions of methodological approaches in published research, or overall systems directed at 'producing' outcomes—from papers to careers. Other issues concern research ethics and addressing sensitive topics in the project results. For example, the expert from the interview I1 reflects on an ethical and practical challenge they faced while attempting to contextualise research results in detailed descriptions about study participants. I1 could not provide such descriptions due to the study participants' concerns about privacy and anonymity. The expert "tried to anonymise all the data" (pos. 18) despite the loss of the contextualisation. To balance the research interests and the study participants' concerns, the expert conducted an overt observation and additional digital ethnographic work in order to complement their findings with publicly available information about the community their worked with. Similarly, I13 also elaborates on the choices about research ethics and anonymity that had to be made for reporting study results at the cost of certain—illustrative—stories that, being told, would have exposed the identities of the study participants.

While these frustrations and challenges concern not only datafication scholarship, but academic research and writing in general, the experts also identified several issues specific for writing and talking about datafication processes. For example, I24 mentioned that writing a book takes so much time that particularly in studies of datafication, the material can become outdated by the time of publishing. For I8, preserving anonymity of study participants and other stakeholders also meant access restrictions to relevant materials (e.g. charts, spreadsheets). Even when access to such materials for researchers was possible, these findings could not be used directly for public presentation of results.

> "So, often, we've ended up talking about those kinds of artifacts, kind of the documentary evidence. We've ended up just talking about and rather than actually showing, which can be

> difficult because it's quite hard to visualize really the extent-, some of the data. And I think it would be quite useful sometimes actually to be able to show, you know, in a conference paper or something, it'd be quite useful to be able to show the spreadsheet, but it's difficult" (I8, Pos. 9).

Further ethical challenge pointed out by I8 concerns trust of study participants not from the perspective of preserving their anonymity, but the representation of study participants in the research findings. Sometimes, so I8, scholars end up criticising their study participants in their publications, that can also undermine mutual trust they built over the course of the research project. The important academic expertise in critical research on datafication, then, is to balance trust and critique. The expert I22, also elaborating on writing about sensitive topics, explains how stating one's own positions towards the topic explicitly in the writing allows to produce more situated accounts about the studied datafication processes.

> "But I remember that was on my mind, what's going to happen here? Especially seeing all the colleagues being trolled for doing sort of equality and diversity work. I think when we were writing the [anonymised] paper, we were really, really careful to maintain, and I hope this was successful, but to maintain our respect for the participants, even when we thought that they were being really sexist, and it really bothered us" (I22, Pos. 17).

Also, I22 reports about personal experiences related to research topic and "awful" (pos. 9) feelings brought forth by contradictions between personal views and those of the study participants. Alongside these feelings go also worries about potentially being insulted by the study participants for a critical perspective on their practices and concerns to be prohibited to use their (publicly available) materials in publications (see also the above quote by I22).

Other ethical challenges address concerns or restrictions which further stakeholders such as developers and social media platforms put on research processes. In one of the projects discussed by I14, ethics has been considered from the perspective of privacy by design, which is, however, criticised by the expert amid the lack of a feedback loop from the users in such approaches. Rather, the aim of the I14's project was to give their study participants tools to (at least try to) exercise their privacy rights or learn more about them. I28 provides an example of their work with data crawlers that is usually not allowed by social media platforms but is required in order to understand, deconstruct their algorithms and hold the companies to account. Overall, in datafication research, making decisions about ethics and anonymity does not only concern the trust placed into the scholars by the study participants, but also the broader imaginaries of technologies and data. Sometimes it may lead to presenting technologies and data as rather neutral, as all the internal struggles, conflicts, and ambivalences are being written out in order to preserve the anonymity and integrity of the study participants.

In sum, some of the topics discussed in this section such as interdisciplinarity and consortium projects, academic writing and publishing, or research ethics do not concern solely datafication scholarship. Several datafication-related issues such as politics of data, often lacking access to materials, stakeholders, and research situations, and an interdisciplinary character of datafication processes provide some examples for the specifics of datafication research politics. By attending to these issues from the perspectives of the interviewed datafication experts, I aimed to illustrate with this section, why data scholars explore datafication processes before exploring how they conduct their research. This section illuminated the decision-making and affective processes required for a research project on datafication to be conducted and the manifold of challenges datafication researchers face. In hindsight of these decision-making practices, related challenges, and care they require, the question, what people, research instruments, artifacts, and elements of datafication processes researchers assemble in their empirical work and how these assemblages help understand datafication better, becomes central. The next three sections of this chapter attend to this question in more detail. Bei outlining various elements of methods assemblages, I show when and how they are performatively enacted and produce certain concepts about datafication.

6.2 Exploring encounters with data representations

The first methods assemblage discussed here allows scholars to study how various people encounter data representations—derivatives produced from data analysis. In this section, I present the results of my analysis drawing on the definition of a methods assemblage developed in chapter 3. According to this, a methods assemblage is an enactment of the relations between the researchers including their positionings in the field, the researched actors, sites of practice, and specific research procedures and techniques. The subsections of section 6.2 refer to these elements of a methods assemblage. First subsection introduces data representations as a core of the methods assemblage. Second subsection primarily addresses kinds of knowledges sought by exploring encounters with data representations; this is followed by a discussion of the relevant research procedures and sites of practice. I draw on a heterogenous sample of empirical projects on datafication discussed in the expert interviews and provide anonymised accounts of these: researched actors, therefore, are mentioned throughout each subsection. This section is concluded with a discussion of how the methods assemblage for exploring encounters with data representations sometimes overlaps with other methods assemblages.

6.2.1 Encountering data representations

Scholars applying methods assemblage of analysing encounters with data representations have different takes on the question of what data representations are, depending on the disciplinary affiliation or research questions of each research project. For example, for some such data representations can be understood as statistical data in form of lists, tables, and various graphic visualisations such as graphs or diagrams (I2, I5). Some scholars attend social media and app use data visualisations (e.g. I12, I14, I23) while others are interested in representations of material, physical objects which become datafied through information systems and data infrastructures (I17, I4). Following the expert interviews, I do not focus here on the methods or tools with which data representations have been created. Rather, I discuss in this section how various forms in which digital data are rendered accessible to human actors (in contrast to machine-readable formats of digital data) being negotiated, created, encountered, and related to by these human actors. Data representations should not be seen as mere visualisations of certain datasets with the help of graphs, lists, or tables; rather, data representations can be analysed as a way to create relations between digital data and actual, lived realities. So, I17 notices in their analysis of data representations (such as numbers meaning to represent certain populations) how even welfare projects use data representations as political instruments for securing and maintaining power positions in relation to other state or nongovernmental actors. I17 was particularly interested in the ways their study participants attributed value to data representations and physical, material objects these data meant to represent (in the case studied by I17, e.g. housing or household items). I17 addresses this value attribution as "ontological relevance" (pos. 13). For example, I17 observed situations where their study participants attempted to 'game the system' and tweak the numbers in order to create such a representation of their lives that it would allow them to better their living conditions (e.g. receive different housing).

Attending to the perceptions and practices upon encountering data representations, I17 explored how study participants translated encountered data representations back into objects and items these data meant to represent and subsequently acted upon their interpretation of data representations, e.g. by tweaking the numbers. In relation to translation taking place in encounters with data representations, I29 points out a methodological challenge for researchers, addressing how researchers and practitioners might understand, perceive, and experience differently what counts as a data representation. They argue that "there's always a translation exercise that takes place when you're using data or even understanding datafication as a phenomenon" (pos. 17). Particularly when study participants deem data representations irrelevant or inaccurate and choose

not to engage with them further, not to act upon them, researchers face a challenge for their interpretive work. As discussed in chapter 2 of my thesis, according to some critical scholars, lacking attention to agency of study participants, especially laypeople, pictures them as ignorant and lacking agency (e.g. Dencik, 2020; Livingstone, 2019). Exploring how people encounter data representations allows to turn away from such a simplified view and investigate in more detail the processes of translation through which people re-situate data representations in their lived realities. So, when talking about encounters with data in such situations, these can be seen as personal encounters, as an understanding of oneself as data and trying to match what is important to oneself to be datafied in the real world. Even when study participants do not take any action upon encountering data representations, they exert agency by re-situating the encounter in accordance with their knowledge, emotional state, and goals.

Respective to the various kinds of data representations and empirical sites of practice, actors encounter data representations differently. For example, in the educational domain, educators encounter data representations in their professional practices in form of national or local test results, rankings of their educational institutions, their students, and themselves. Other, laypeople encounter various data representations in their everyday lives, producing these themselves as part of their self-quantification practices, or encountering data representations produced by other actors, for example while reading a newspaper with a data visualisation in it. In other cases, researchers purposefully create encounters of their study participants with data representations (e.g. I12, I14). In these encounters with data representations, datafication processes and their implications are being enacted. Attending to encounters with data representations analytically, therefore, allows to study the enactment of datafication processes by people who are not negotiating data and their representations during their development, but rather re-interpret these in their everyday lives.

6.2.2 Understanding lived experiences

To understand how data representations are encountered, datafication researchers investigate the actors involved, their perceptions, positions, and responsibilities in regard to processing or interpreting the data representations, who is best informed about the datafication processes enacted in these representations, and how analysed data representations are situated (I5). Central to the methods assemblage of exploring encounters with data representations is not an analysis of the data representations themselves, but rather actors' experiences and perceptions of these. I12 elaborates on the goals of one of their research projects illustrating such an application of the methods assemblage.

> "So we didn't want to analyse what they do on their devices, how much they use for what and that data was not directly available for us. […] well, at least for me, that was not the focus, but I wanted to hear how people reflect when they encounter their own visualisations" (I12, Pos. 5).

In the projects in which methods assemblages of exploring encounters with data representations were enacted, understanding lived experiences of people in their everyday or work lives has been the central research interest (e.g. I1, I2, I3, I12, I17, I18, I29). Attending to the lived experiences of the study participants analytically in addition to or instead of an analysis of data representations themselves can make visible the discrepancies between lived experiences and data which are meant to represent these experiences (e.g. in cases of tracking data stating something different than what study participants perceived they did, such as reported by I12, I23). To understand perceptions of and experiences with data representations, qualitative research is usually conducted. The scholars aim to understand the enactment of datafication processes and/or their implications on various actors. With the help of qualitative research techniques, perceptions such as thoughts, feelings, and attitudes towards datafication, as well as lived experiences including practices and activities performed during encounters with data representations are examined. Different experts defined

lived experiences differently. For example, for I1 these comprise of sayings and doings of their study participants, while I3 give priorities to the practices—what study participants actually do as they encounter data representations. For I2, lived experiences can be understood as "the discourses, the impressions, and how these discourses were reflected in the practices" (pos. 15). For I12, who has been inspired by Helen Kennedy's work, lived experiences particularly address the kinds of feelings evoked by various datafication processes in the everyday lives of study participants. For I5, exploring lived experiences means understanding how educational institutions "engage with these different forms of data" (pos. 29). I8 describes their focus on lived experiences as an exploration of relationships through ethnographically inspired research designs: "the kind of ethnographic style observation of trying to understand the relationships the within a setting" (pos. 29). In I8's projects such ethnographic observation is accompanied by interviews to provide deeper understandings of these relationships.

Analysis of expert interviews suggests that datafication researchers attend to experiences with and perceptions of data representations together. For example, I1 explains how they combined ethnographic observation with interviews to capture both.

> "[T]hat's also why I chose the qualitative approach, […] because I was interested in reconstructing the perspective of the people. So, that's why also I conducted the qualitative interview. Not only seeing what's happening through the […] observation, but I'm interested in, what do these people actually think? […] What is their purpose and what are their aims?" (I1, pos. 15).

In an exploration of encounters with data representations, lived experiences and perceptions collapse together with imaginaries of datafication processes as study participants encounter data representations in their everyday (e.g. I18) or working lives (e.g. I11, I12). Researchers' analytical attention is directed at identifying lived experiences in encounters with data representations and what ideas, feelings, imaginaries, and values are guiding the study participants' perception of these representations. During the encounters with data representations, they are being enacted as such and acted upon. Attending to representations of these diverse kinds of data in the everyday and professional lives of individuals and collectives allows scholars to explore the performative role data play in shaping societies. Studying encounters with data representations means that scholars focus on performative characteristics of data and their representations and how individuals or collectives act upon such data representations.

Some of the reported examples illustrate how perceptions of and experiences with data representations and the datafication processes they enact are guided by the study participants' motivations and interests. For example, I12 remembers how in one of their projects, they expected study participants to reflect critically on the data extraction by big tech corporations when facing visualisations of their app use data. The study participants, however, preferred to view these visualisations as a call to action (e.g. to change one's behaviour), as a way to learn more about themselves, or as an affirmation of their practices. These individual preferences and interests drove the study participants to join the research project in the first place, in contrast to the researcher's aim in empowering critical reflection of the data extraction. If data representations respond to the actors' interests, there is also some generative power of datafication processes and productive transformations following these (e.g. I3, I14, I25). So, I14 elaborating on some of their research projects, argues that datafication research should turn attention to the ways in which stakeholders benefit from encountering their data representations.

> "But I think we also need conversations around what are possibilities, for users -, for citizens, if they have access to their own data, because this is something that they don't normally -, that we don't normally have" (I14, pos. 5).

An encounter with data representations created by research team may be a way to make these productive, empowering implications explicit, while particularly participatory projects lend themselves for such research interventions (e.g. I14).

I18, in contrast, questions the widely applied focus on perception of datafication processes as knowledge of these processes. They question the extent to which datafication processes are knowable at all, even for themselves or other datafication experts because of complexities of the phenomena, limited access to information, and continuous flux of datafication processes: "[U]nderstanding [datafication] somewhat is about the best that I could manage" (I18, pos. 13). I26 makes a similar point wondering about the meaning of the concept of datafication for various actors and questioning whether and how datafication processes can be understood at all.

> "Like people are just talking about datafication like as something out there, like an abstract thing sometimes. If it's so hard to pin down what it is that we're actually talking about, what does that mean for our understanding of datafication?" (I26, pos. 19).

According to I18 and I22, focusing specifically on data representations such as visualisations allows to grasp datafication processes in greater detail, rather than attempt to describe perceptions of datafication in general. Visualisations, while having their own aesthetics and rules (e.g. I22), can be more easily re-situated, for example by exploring how they were created, and evoke emotional, affective reactions. I12 recounts how in the research project they conducted, study participants experienced different kinds of feelings upon an encounter with data representations, ranging from anxiety to satisfaction. I22 provides another example on why feelings matter in an analysis of encounters with data representations by turning attention to why actors encounter data representations in their everyday lives.

> "You can't actually get people interested in stuff they don't care about. Right? So, there was a sense that data was just being consumed on a sort of 'need to know' basis. I mean, just, "I need to do this for work," or, "This is part of the app that I'm using to track my exercise." But for the most part, that was by the by, and I don't think people really associated that with data in the way that we were talking about it on the project." (I22, pos. 15).

This quote echoes the proposition of the expert I18 to study in detail a small number of datafication processes as enacted in encounters with data representations. Such detailed attention can allow differentiation between a variety of experiences that enable manifold ways in which encounters with datafication are perceived and known by people. In research projects conducted by I12 and I22, an encounter with data representations was orchestrated by the research team and used as a prompt to elicit how study participants make sense of data representations emotionally and cognitively. Methodologically, an encounter with data representations not only works as a prompt in an empirical study, but also allows researchers a better understanding of why certain data representations are valued by actors and not others, how these actors interpret these representations, and how they act upon such an encounter. An analytical switch from such interpretation as cognition to further forms of knowing about datafication based on encounters with data representations also means for some interviewed experts (e.g. I18) moving from iconic examples of such data representations to more mundane ones, those deeper embedded in the lived experiences of actors, in their everyday and professional lives. A methods assemblage of exploring encounters with data representations renders lived experiences of people living in datafied society visible for research, particularly when research aims at understanding everyday personal or professional lives of various actors.

In research applying methods assemblages of exploring encounters with data representations, study participants are not the only ones who encounter data representations. Particularly in participatory projects in which tools for developing data representations or representations themselves are being developed collectively with the study participants, researchers do not necessarily expect certain results and also encounter the representations for the first time. For example, I14 reports on learning "as much from, you know, our own experiences, but also from the eyes of our participants who were also seeing these things" (pos. 9). Another example is I27, for whom the whole research projects described in the expert interview started from personal experience with datafication processes and the discrepancies between data representations and what

these meant to represent. Other experts also tried out different research tools themselves in order to re-enact themselves the encounters with data representations, which the participants of their studies had. For example, I12 recounts their experiences of working with research tools study participants used to create visualisations of their app use as exhaustive and anxiety-provoking (pos. 26).

> "Then having to be using the Android phone and being exposed to all the data collection that Google is doing, that was kind of eye-opening for me. All the notifications linked with my location like - Would you like to recommend this Museum or this restaurant? I hadn't known that it is there because I was using the iPhone. I also realised how exhaustive it was. So, when the apps were tracking a lot of things, it was just too much to digest so I didn't even-, in the position of being a researcher, I didn't find the time and motivation to go through everything. Especially the kind of log that had a scale that did it by second." (I12, pos. 26).

I12 also notices that they preferred certain data visualisations over others. I7 also tried out tools their study participants were using to create and encounter data visualisations, similarly, finding these practices personally annoying.

6.2.3 How are encounters with data representations studied?

Encounters with data representations that both study participants and researchers make take place as a part of various research designs. Particularly participatory research designs, however, are core to the methods assemblage of exploring encounters with data representations. Conducting participatory research, scholars create and/or observe situations of encounters with data representations in order to empower them in understanding their data representations and becoming actionable upon these (e.g. I10, I14, also applying participatory methods I18). Within some of the participatory research projects, data representations (e.g. I12) or software that allows creating new data representations (e.g. I11, I14) can be developed together with the study participants. Through the joint development or at least requirements analysis for the development of such software, experiences, values, and priorities of the study participants in regard to the expected data representations can be understood better and situated in other, related practices and their individual and collective backgrounds/histories. One of the projects discussed in the interview with I14 provides the most illustrative example of application of participatory approaches within the methods assemblage of studying data representations. Here, members of the local community with whom scholars worked together on the project not only functioned as informants, but actively shaped conversations and research findings. Based on that, I14 and their colleagues developed an understanding of datafication processes taking into consideration personal experiences and "punctuated moments that they [meaning, study participants] thought were also somewhat strange" (I14, pos. 9). Another example of a participatory research project is work elaborated on by I3, who co-organised meetings, events, and produced results relevant for the studied community together with the community members. Most examples of research applying methods assemblage of exploring encounters with data representations focuses on individual people as study participants. Some of the above examples, however, such as in projects conducted by I1 or I14, described their research projects as being aimed at studying certain collectives of people (e.g. activists or local communities). I3 also reports how their study participants addressed issues of collective control and use of data in their practices. Thus, lived experiences of the study participants do not only describe their individual perspective and values concerning datafication processes, but can also represent collective perspectives and experiences.

Based on the expert interviews, ethnographic methods (ethnography including virtual one; observations, ethnographic interviews; I1, I3, I17), discourse analysis (I2, I22), and participatory research methods (I3, I5, I14) can be identified as central for the methods assemblage. For I1, an exploration of lived experiences is achieved with ethnographic methods to study practices (doings)

and perceptions of datafication processes (sayings), as well as physical and virtual materials their study participants produce, encounter, and operate with. I1 mentions that applying virtual ethnography "made it easier because it was written information and much easier to get it" (pos. 18). In contrast, observations conducted during actual meetings of study participants allowed to capture the materialities of datafication processes (technologies as well as materials made by the study participants). I26 also attends to the materialities of datafication processes in general and data representations in particular through their ethnographic work. For them, spending time at the empirical site of practice (public authorities and other organisations), taking pictures and making fieldnotes allows to capture the "feel of the place" (I26, pos. 9) for the later analysis. For I11, observations primarily allowed to understand study participants' practices of engaging with professional software and data representations this software allows to create. In addition to observations, some experts used methods of self-observation by the study participants (e.g. I12) to mitigate the limits of digital tools such as tracking apps that do not work cross-platform. Acknowledging the role of data politics in studies of datafication processes, I17 views ethnographic methods as more suitable as these allow staying sensitive towards power relations arising through datafication processes and performed/enacted in data representations. While most interviewed experts recounted the advantages of conducting qualitative, interpretive research, I29 provides a contrasting example in that application of creative, future-oriented methods constrained an analysis of lived experiences: "too much speculation about what the [anonymised] person might have meant" (pos. 3).

Alongside with ethnographies and interviews, few interviewed experts also conducted 'desk research' or discourse analysis of various literatures such as policy documents, textual and visual materials produced by the study participants in each of the reported projects (e.g. I1, I2, I5, I22). In the policy research conducted by I2, policy documents were considered as human-made: "they express ideas that are outside kind of the borders of where the document was written" (pos. 17). As the next sections of this thesis chapter will show, such perspective on the textual materials produced at the empirical sites of practice is also shared by scholars who applied the methods assemblages of reconstructing datafied regimes and for tracing dynamics of data infrastructures. For two latter methods assemblages (policy) documentation is often one of few other resources available for research that allow an insight into the ways datafied regimes and infrastructures are shaped and how/what enables these. In contrast, datafication scholars enacting methods assemblages of studying encounters with data representations view textual materials as artifacts[21] used by the study participants, while their focus is on the implications and effects these artifacts have. Interestingly, several of my interview partners did not agree with the word "artifact". For example, I19 mentions that they "don't like the term 'artifact' as much, even though you… you can use it" (pos. 7). Another example provides I22, who examined both text and data representations such as visualisations on the webpages used by their study participants. These discursive and visual analyses served as an addition to other methods (web crawling) to understand how and when study participants encounter data representations (I22). I5 conducted document analysis of policy documents and other materials used in educational institutions to situate findings acquired with the help of other methods. I6 also used a combination of various methods to situate their findings. Their research design, however developed out of necessity to complement desk research of relevant public authorities' websites and documents, which did not provide information sufficient for research goals, with other methods.

One of the interviewed experts who used to conduct surveys in some of their projects to study encounters with data representations, however, themselves offer critique of the method (I18).

[21] In my thesis, I use the term 'artifact' in order to delineate the materials (texts, artwork, etc.) produced by study participants in the projects reported by expert interviews, from data representations as an analytical categorisation developed in my thesis for the methods assemblage discussed in this section.

They recount how asking about 'understandings' of datafication is nearly impossible for anyone, including datafication scholars themselves: "And we are, you know, leading datafication experts going, what's the right answer to that? You know, we don't know the right answer to that" (pos. 13). Despite the complexities of datafication processes, data politics, and lack of access to data or algorithms, surveys often include questions that would require complex answers about the extent to that survey participants understand certain aspects of datafication. I18 also laments the "obsession with knowledge" that they observe in survey designs and suggest attending to other forms of knowledge, such as affect which laypersons can more easily assess and report in surveys. Making similar argument, I8 underscores the limitations of survey-based research:

> "Well, the survey data, some of it is completely statistical, so that's kind of, you know, it has its real limitations in my view in terms of explanations, but it's there as background. We also collected quite a lot of written data as part of the surveys" (pos. 15).

Rather, in I8's research practice, surveys are used to gather quantitative research data important for developing more generalised findings about broader societal issues and their political outreach.

Overall, in most projects that I identified as enacting the methods assemblage of studying encounters with datafication, explorative, qualitative studies were conducted. Other scholars also developed or used various kinds of (research) software to create data representations and encounters with them (app data trackers, crawlers). For example, in the projects led by I11, workshops with study participants were conducted to understand their requirements for developing a certain software that would help them create data visualisations. I7 and I12 used various data trackers together with their study participants on different kinds of devices (laptops, mobile phones, smart watches). I22 also used crawlers to explore encounters with data representations. In the course of their project, however, they could not explain certain bias in the crawling results. This example demonstrates how usage of research software and (in this case, computational) methods is performative to the findings and concepts they produce. Regarding methods of data analysis, only few experts explained explicitly how they conducted analysis, while most of them applied methods similar to or inspired by grounded theory (e.g. I3).

For some of the interviewed experts who enacted methods assemblage of studying encounters with data representations, datafication processes were not an initial research interest but rather developed as an empirical finding (e.g. I2, I8, I17, I18). Studying lived experiences of various actors foregrounded an important role various kinds of data representations played for these actors everyday or professional lives. Particularly those experts who study implications of policies in different societal domains or are interested in practices of social media use understood datafication as one of the many processes enacted at the empirical sites of practice. During their projects, however, the timeliness and importance of datafication processes became a core finding. For example, in I2's research project, datafication was sometimes seen as a problem addressed by the study participants. As data representations are also social and the results of negotiations of various actors involved in their creation, I10 underscores that besides their function to represent reality, data representations showcase other intents. Similarly, for I12, one of their main research interests were motivations, values, and interests that guide people in the interpretation of their personal data. An analytical attention to the perceptions of and attitudes towards datafication processes enacted through data representations does not always allow scholars to follow up on what kinds of actions people take after encountering data representations, thus bracketing their agency from the analysis and subsequent concepts of datafication.

6.2.4 Encounters with data representations and other methods assemblages

Scholars enacting several methods assemblages at the same time, in contrast, address not solely lived experiences of datafication processes, but also their social, political, and other implications including kinds of agency and actionability various actors acquire upon encountering data representations. For example, in their project that comprises elements of all three methods

assemblages, I17 could observe how for some of their study participants—political actors—developing and operating with various data representations was a strategic task required to achieve political goals and accumulate political power. Similarly, I5 explored how educators acquire new kinds of actionability upon encountering educational data representations and, at the same time, grapple with the expectations to act upon these data in certain ways. Examining educators encounters with data representations and their subsequent actions, I5 could identify how alongside data literacy, professional skills, personal as well as professional values, and goals are central for the kinds of actions educators can take upon encountering data representations such as students' achievement data rankings. They recount how educators

> "are assisting one another with professional conversations and professional engagement with one another, to try to improve their own understandings […], how they are going about their teaching, and how they are assessing students" (I5, pos. 31).

Although in their analysis I5 observed some workarounds educators develop, they argue showcasing these in research downplays the risks of datafication and standardisation in education so no systemic change can take place.

> "And the tension I think around that is that, you know, there is now argument that this standardisation practices are very prominent, and it can have really quite significant deleterious effects. And so, you run the risk, I suppose of sort of downplaying just how problematic some of those things can be, as you are trying to identify ways in which [educators] have to try to be much more productive in their engagement with students' work. And you know, how they try to work around some of the problematic aspects of these standardisation processes. But I guess the question is, you know, should [educators] have to do that or should systemic change, a reform, is something which mitigate against some of those problems, the standardisation doesn't actually contribute to that." (I5, pos. 31).

The example provided by I5 underscores how educators encountering data representations in their everyday professional lives have leeway for acting upon such representations in ways not expected and not prescribed by educational authorities. These actions, however, remain an exemption and picturing these analytically as educators' agency hinders rather than promotes more systemic changes that would actually grant educators more actionability upon encountering educational data representations. In another example of research on datafication in education, I29 and their colleagues could observe how during some periods of time—during tests based on which educational data representations such as ranking should be developed—educators' actions were directed solely at these data representations. I29, therefore, questions the relation between data representations created through these tests and students' learning these data are meant to represent. At the same time, I29 and their colleagues observed how such data representations—education rankings—while published provided only decontextualised accounts and did not allow much local actionability for involved stakeholders.

> "There were social consequences of the instrument, because during the testing periods, when the […] tests were held, schools are wildly preoccupied with that test at the expense of other teaching and learning. Okay? There's that. But here's the most problematic aspect, [there] was a [political actor who] made the decision to present that data publicly on a website […]. And pretty much the minute that that data was made publicly available, it was used to create league tables and schools entered into a fiercely competitive- Which sort of rolls all the problems with the use of data into one. The type of data they're generating is problematic. It's not used for its original purpose, and then it's having these warped side effects that just intensified the competition." (I29, pos. 15).

For I29, to allow informed agency upon encountering certain data representations, these representations need to be re-situated. I7 also questions whether 'having' data about something and encountering their representations (e.g. visualisations of app tracking data) creates situations in which individuals are expected to act upon these data in a normalised way. For example, if data tracking visualisations can provide more control over what tracking data visualisations illustrate as

excessive behaviour. For I7, a data representation of "certain parameters of the self, especially of the body, but also of the behaviour" (pos. 17) therefore, produce normativities. They recount how one of their study participants explained that they "can never ever in [their] life eat sweets without thinking all the time what it means. That [they] have to go on a run in the park" (I7, pos. 29). In this example, different kinds of knowledges about datafication are highlighted. On the one hand, the expert I7 was interested in individuals' tracking activities and their lived experiences of encountering visualisations of their tracking. This is central research interest pursued by scholars enacting methods assemblages of exploring encounters with data representations. On the other hand, questioning the extent to which these tracking data visualisations re-produce normativity, and who negotiated what kind of action needs to be taken upon consuming certain foods, the expert draws attention to the kinds of knowledges primarily sought by applying the methods assemblage of reconstructing datafied regimes, discussed in chapter 6.4

In sum, methods assemblages of exploring encounters with data representations primarily serve to understand lived experiences of individual users who encounter data representations in their everyday or professional practices, while some experts studied how collective actors work with various kinds of data representations in professional contexts. Lived experiences in their different forms such as perceptions and imaginaries of datafication processes, practices, and feelings about these can be studied. By attending to encounters with data representations, scholars explore how their study participants—often 'ordinary', laypersons—translate and re-situate encountered data representations in their everyday and professional lives. While some datafication researchers are primarily interested in perceptions and affectivities in relation to encountered data representations, others foreground what people do when they encounter data representations and how they become actionable or not upon such an encounter. Altogether, methods assemblages for exploring encounters with data representations allow researchers to study how datafication processes are being enacted in lived experiences of various people, often in mundane, everyday situations. Particularly through participatory approaches, researchers exploring encounters with data representations learn not only about their study participants, but also together with them how datafication processes are being enacted as people translate and re-situate data representations in their lived realities. Finally, the interviewed scholars foreground actionability upon encounters with data representations, while some of the experts specifically draw attention to positive, generative examples of such actions, also acknowledging challenges of such an enthusiastic representation of positive examples of agency. The purpose of this section was to elucidate how methods assemblage of encountering data representations is ordering various elements of research processes and empirical sites of practice. This section showed that datafication scholars enacting this methods assemblage are primarily interested in the lived experiences relevant for the studies datafication processes. These lived experiences also encompass actionability upon encountering certain data representations. Foregrounding encounters with data representations as a methodological approach to exploring datafication processes should demystify data as something that only experts can talk about and relate to. Instead, exploring encounters with data representations shows the extent to which laypersons are related to data and how in their everyday personal and professional lives they re-enact datafication processes, translating and re-situating data representations into what is important for them.

6.3 Tracing dynamics of data infrastructures

Each of the following subsections presenting my findings on the methods assemblage for tracing dynamics of data infrastructures refers to the central elements of the methods assemblages. In the first subsection, data infrastructures are discussed in relation to the kinds of knowledges researchers seek as they enact this methods assemblage. In the following subsection, I elucidate research procedures and sites of practice relevant for tracing dynamics of data infrastructures according to

my sample. The final subsection is dedicated to overlaps between all the identified methods assemblages with the methods assemblage for tracing dynamics of data infrastructures.

6.3.1 Dynamics of data infrastructures

Scholars who enacted in their studies methods assemblage of tracing dynamics of data infrastructures are concerned with the socio-economic implications of data infrastructures (e.g. I4, I17); others primarily attended to historical developments of infrastructure governance and their political implications (e.g. I4, I15, I16, I19). Interest in various social and political implications of data infrastructures enabling datafication processes is closely intertwined with and extends on scholars' attention to the lived experiences of the impacted people. In contrast to the methods assemblage for exploring encounters with data representations which foregrounds lived experiences of individual people upon such encounters, scholars enacting methods assemblage of tracing data infrastructures study the implications of these infrastructures on lived realities of various groups of people, e.g. living in one region. Unlike in research projects exploring encounters with data representations, for scholars tracing dynamics of data infrastructures, these infrastructures and data themselves are important actors in enacting datafication processes and are addressed analytically (e.g. I4, I14, I15, I17, I21). For an understanding of data infrastructures, a research project mentioned by one of the interviewed experts, I12, is illustrative. In the interview, I12 elaborated on the "*Anatomy of AI*" (Crawford & Joler, 2018) project, highlighting how it helps to grasp the materialities, heterogeneities, and complexities of infrastructures involved in enacting datafication processes. Most experts enacting methods assemblage of tracing dynamics of data infrastructures report about their research projects as being based on tracing some big political infrastructural projects brought into life by governments and/or other public authorities (I4, I13, I14, I17). In such projects, understanding of infrastructures includes both building physical, material infrastructures connecting different places through facilities, cables, pipes, sensors, data, and their categorisations (I30) as well as staff recruitment and training in operating these.

The term 'infrastructure', similar to 'datafication', poses a particular challenge for researchers, as it can be understood both as an empirical phenomenon and a theoretical concept (e.g. "datastructuring", see e.g. Flyverbom & Murray, 2018). In the empirical studies of data infrastructures, therefore, the experts expand on different understandings of what data infrastructures are and how these are involved in enacting datafication processes. For example, I19 reflects on infrastructures conceptually as being a dynamic and a stable entity at the same time, while different aspects of it come to the fore in different situations[22].

> "I find this approach quite innovative, to say, I bring together empirically and conceptually this well-defined idea that there is some dynamic and power of the infrastructures and, at the same time, I address this non-determination, contextuality and try to understand how they work together" (I19, pos. 23).

In another example, I17 initially not aiming to study datafication processes, began their research by engaging with technological infrastructures (outside of Europe) and the role these play in the everyday lives of people dependent on these infrastructures. During ethnographic observations of the study participants' lived experiences, I17 identified data as a crucial actor mediating between public authorities who provided the infrastructures and the people dependent on the latter. This agentic understanding of data infrastructures, then, shifts attention from the actual practices and encounters with data representations to the social, economic, political, and cultural implications of these encounters. At the same time, I17's research project also provides an example of a study applying elements of different methods assemblages: both assemblages for exploring encounters with data representations and for tracing dynamics of data infrastructures. I19, in contrast, primarily attends to collective actors (e.g. public administrations), their relations to each other established

[22] This definition builds on Star and Ruhleder (1994, 1996).

through data infrastructures, and their spatial proximity as they explore the implications of these relations for politics and governance.

Another challenge of data infrastructures is that these are understood as complex and distributed. Thus, it is difficult for researchers interested in their social, political, cultural, and economic implications to single out the 'cause and effect' of data infrastructures or some of their elements (I19) or to define infrastructures in the first place (I24). Instead, interviewed datafication researchers attend to the "changing processes of decision-making or type of logic or ways of understanding people" (I16, pos. 19) to understand how datafication processes are enacted through data infrastructures. Being distributed across various physical and virtual spaces and actors, data infrastructures enact different "kind[s] of datafication" (I16, pos. 29), depending on the analytical perspective of the scholar or practitioner, the studied aspects of infrastructures, or areas of their empirical application. To situate data infrastructures analytically, methodological cuts (Barad, 2007) are required to make the infrastructures less complex and more "graspable" (e.g. I19, pos. 11). Such methodological cuts allow moving from infrastructures to actors and their practices, necessary to understand the enactment of datafication processes through data infrastructures. Making cuts, however, responds to the analytical framework of a researcher (and their teams) and not necessarily to that of the practitioners involved in the design, development, or maintenance of data infrastructures. Like researchers, practitioners, too, do not have an ultimate overview over their data infrastructures due to their distributedness and complexity. I27 reflects on that stating that even researchers and developers of data infrastructures are not able to fully comprehend their implications:

> "I don't know that we ourselves know what the right blend of expertise is for crafting, analysing, and interpreting the data, that a lot of these infrastructures are capable of creating" (pos. 12).

Therefore, even when individual actors are reflective about their roles in the creating of data infrastructures and the implications the latter have on further stakeholders, the implications of the whole infrastructure are difficult to estimate. Paying attention to such individual perspectives of actors involved in design and development of data infrastructures allows some experts (e.g. I19) to build up awareness and tacit knowledge and trace the (epistemological, ideological) background of these infrastructures. According to I29, however, studying a single part of data infrastructures in isolation (in this example, infrastructures of data analytics), is not necessarily covering the social, cultural, political aspects of the studied phenomenon. For I29 these aspects are of utmost importance for understanding implications of datafication processes on society.

> "I have colleagues who are deeply immersed and very, very good at learning analytics. I'm not so sure they're aware of the pedagogical political, economic consequences of what that work's doing. And so, one of the problems is you study a thing in isolation and generate data about it. But if you're not aware of how that data might be used for good or bad, and otherwise, you need to be aware of these different implications." (I29, pos. 17).

Some of the interviewed datafication researchers, therefore, trace historical developments of data infrastructures back to their infrastructural, political, and social predecessors (e.g. I16) and imaginaries (e.g. through works of fiction, mentioned by I15) to re-situate these in the lived realities of the involved actors. These experts emphasise the analytical ability to situate data infrastructures and their historical development over the knowledge of technical, computational parts of the data infrastructures and related datafication processes. For example, I4 and I5 explain their approaches to research designs for studies on data infrastructures through document analysis as exploring the work 'behind' data infrastructures.

> "For example, some people ask me: „Oh, can you tell me about the platform?" No, I cannot because I am not a techie. I am not-, I am someone that studies how technology (impacts) people. So, we see it at the interface between the technology and the person. […] So, we start from there. And then my inquiry, I mean, it is normally when it is not Covid, is normally

> interview based. And the completion and participation based like being in the places where we encounter them, where people encounter technology. And then to trace back the mechanisms, the processes behind those encounters, to do so, I use a lot of historical secondary data. So, archival data, and the government reports, everything really. So, methodologically, again I am a qualitative researcher […], so sit down with the technology, with the people. But yes, to trace the historical process is what we find most important." (I4, pos. 19).

In some cases, tracing historical predecessors of current datafication processes reveals that data infrastructures do not bring forth that much of a change in terms of practices of individual actors, although what is changing is the scalability of these practices (I9).

> "So, as I said before, yeah, how little is actually new about it. If you actually look at what people do. You always think, [anonymised site of practice] is somehow cool, new… So, tech is something completely new, but there have been so many predecessors before which were very, very similar." (I9, pos. 13).

Within my sample of expert interviews, research projects tracing dynamics of data infrastructures investigated such infrastructures developed by governments and public actors, while in several cases these data infrastructures were specifically developed to provide services for marginalised communities (e.g. I4, I13, I17). In such government programmes, analysed data infrastructures are used, at least to some extent, to accommodate the needs of marginalised communities such as access to shelter, food, water. Therefore, research projects reported by the experts in my sample here might have had a significant role in constructing my understanding of the methods assemblage of tracing dynamics of data infrastructures. Some of the interviewed datafication researchers explained that they have been invited to accompany the development of such projects analytically either by the public authorities themselves, or by non-governmental organisations (NGOs) also working with the same citizens who should benefit from data infrastructures.

> "[A]nd then suddenly we got-. It was actually a talk that I gave in [a country] and somebody came to me afterwards and said: "Well, you should come and look at our [programme] because there's a really interesting person who joined", so it was a huge amount of serendipity involved as well, but I think probably if you stay in a place long enough and if you hand around the project long enough opportunities will come. So, we were very lucky to get access to this other [programme] who was doing things in a very different way to what we'd seen elsewhere." (I13, pos. 15).

The state and public programmes within which data infrastructures are developed can be seen as means of achieving social change, as I13 puts it (I4 and I17 report about similar experiences):

> "there was an intent being made to radically transform the way that public […] services were being delivered in [country] and computers and data were seen as a very central aspect of this" (I13, pos. 15).

In another example, I17 talks about a project aimed at improving the community life in a certain region, although through data infrastructures, new political and power relations developed, complicating the achievement of the initial goal. The research project described by the expert addressed the gap between set goals and datafied realities, illuminating the "strategic importance" (pos. 3) of data to the actors involved in data infrastructures at the empirical site of practice. Researchers tracing dynamics of data infrastructures, therefore, are mostly focused not on the architecture of the infrastructures, but rather on the implications of introducing data infrastructures to the people and the implications these data infrastructures have on populations, particularly those being excluded from social, economic, and political relations through these data infrastructures (e.g. I4). Apart from formal research techniques, informal sources are vastly used to understand the researched site of practice better. However, due to their relations to particular organisations on site (e.g. NGOs) some scholars explicitly position themselves towards the infrastructures they study

(e.g. offer explicit critique of how infrastructures may be used for governance or particular political goals, e.g. I4).

Research projects enacting methods assemblages of tracing dynamics of data infrastructures are not exclusively directed at understanding the lived experiences of individuals or professional practices of people developing and maintaining such infrastructures. Instead, this methods assemblage makes a turn to the ongoing re-enactment of datafication processes between the actors enabling these processes and actors affected by these. What is dynamic about data infrastructures, therefore, can be summarised in three ways. First, as at least some of interviewed scholars extend on the conception of infrastructures by Star (e.g. Star & Ruhleder, 1996), data infrastructures themselves can be understood as processual rather than stable. Second, infrastructures are developing over time, therefore some of the scholars attend to their temporal dynamics either purposefully, by examining their historical predecessors and future imaginaries, or not purposefully as researchers are confronted with delays and disruptions in the infrastructures' development and maintenance. Finally, some of the interviewed datafication researchers attempt an empirical investigation of what Star and colleagues address as processual: these interviewed experts specifically examine the work of various human and non-human actors required to maintain data infrastructures and for the infrastructures to serve their purpose of moving data from one actor and information system to another. For this, the interviewed scholars enacting the methods assemblage discussed in this section either study the actors involved in the data infrastructures and their relations to each other or focus on the work some of these actors do to keep data (and other things such as material goods) moving. The following sections elaborate on this in more detail.

6.3.2 How are dynamics of data infrastructures traced?

To enact methods assemblage of tracing dynamics of data infrastructures, scholars use various techniques of data collection and analysis. According to the experts, interviews accompanied by observations (e.g. I4, I13, I14, I16, I17, I18, I26) and policy research (e.g. I4, I15) can be identified as central for this methods assemblage. Overall, in most projects that I identified as applying the methods assemblage of tracing dynamics of data infrastructures, either longitudinal studies or multiple projects over a longer period of time were conducted (e.g. I4, I13 have been working on the same data infrastructures projects for many years in a row). Such research is conducted either as part of longitudinal projects, or across multiple projects dedicated to similar topics. Such a long-time research commitment allows scholars to develop and gain tacit—implicit—knowledge about the data infrastructures as being embedded in cultural, political, social, and economic relations at the studied sites of practice (I4, I13, I15, I17). Tacit knowledge acquired through these relations, is required both to situate new findings at the site of practice, and to show stakeholders a level of awareness about their practices required for a productive exchange with these stakeholders. I13 explains how such tacit knowledge allowed them to produce situated knowledges about data infrastructures and to understand how datafication processes are enacted with and through these infrastructures.

> "So I suppose we could have gone there in the early years and we could have whipped up some story about what was going on and described all our data collection methods and all these kind of things and I'm sure we could've come up with some kind of story and we could've made it fit in some kind of thing but it seemed to us that we weren't really getting at very interesting phenomenon and we really had to be hanging around for a longer period time and maybe it was just that we were lucky and maybe it's just that we arrived too early in the process or whatever but I think, the time we spent there certainly wasn't wasted and we got a really good feel for politics for this kind of process and the difficulties and we got a feel for the way in which these public […] institutions are managed on different levels, at the very local levels […] and so on and the kinds of people and the kinds of-." (I13, pos. 19).

Building trusting, sustainable relations with the study participants and other multiple stakeholders such as political and civic actors is in the core of research designs in projects applying methods assemblages of tracing dynamics of data infrastructures. Such relations to the study participants, partnering organisations, further stakeholders, and researchers are crucial, for example, for gaining access to the empirical site of practice, e.g. to people, documents, organisations, and places that are usually closed for researchers. As I13 summarises it, sometimes there is a long way for scholars to become a member of a relevant community and be able to draw on their expertise for developing new research topics regarding datafication processes. Another expert, I15, also working in a country different from their country of residence, reflects on how "understanding the context. So, understanding the country, understanding maybe culture and history" (pos. 11) is also a part of gaining tacit knowledge about the empirical site of practice. Due to the central role of tacit knowledge in tracing the dynamics of data infrastructures within political, social, cultural, and economic contexts, senior researchers face the challenge of building up awareness in the junior team members (I21). Particularly in the domains such as data infrastructures of social media apps and platforms, the personalised use experiences make this task more complicated. At the same, at least in the case reported by I13, a long-term commitment by researchers means accepting possible negative consequences. I13 remembers the feelings of frustration due to the long unproductive phases within a longitudinal research project. The expert explains that "from the methodological point of view, I think the most important thing was sticking with it and staying there over and staying involved over a long period of time" (pos. 15). A long-term commitment to data infrastructures as an object of study also means following the pace of infrastructures development at the site of practice and embracing the disruptions that take place due to changes in policies (e.g. new government that follows a new agenda in regard to the data infrastructure programme), delays in facilities building, or staff training (e.g. I13, I30). I13 also mentions how their career stage (tenured faculty member) allowed them to continue working on the project and allowed waiting over several years for the data infrastructure programme to develop further.

Data infrastructures in the studies reported by the interviewed experts are tightly intertwined with infrastructures for distribution of material goods (e.g. I4, I17). A connection between data infrastructures and material goods has several consequences for the people utilising these infrastructures. First, data acquire new materiality and immediacy as means to acquire goods, similar to what has been discussed in the previous section of this chapter in regard to the data representations' 'ontological relevance' (I17). Second, due to this connection between the data infrastructures and the distribution of goods, political economies of goods distribution and data extraction also get intertwined, while the affected people (e.g. citizens) attempt to situate themselves and their data within both in order to benefit from them (I17). Both I4 and I17 in their analyses of data infrastructures built by public authorities notice how issues of physical access to data infrastructures become political: regulating access of affected actors to the data infrastructures, public authorities and political actors produce inclusions, exclusions, and (in)visibilities. Particularly in programmes within which data infrastructures aim to support citizens, such regulation of access to data infrastructures illuminates implications of datafication processes enacted though these infrastructures on the very notion of citizenship. Then, datafication research tracing dynamics of data infrastructures attends to the different and continuously changing ways in which datafied goods and identities are enacted, disrupted, and come to matter (in comparison to the flows of actual goods, I4). Scholars conducting such datafication research primarily apply ethnographic techniques that are sensitive to material, physical aspects of studied datafication processes.

> "So, actually my main concern has always been with the people. With the people that deal with these entitlements. And I think just very quickly summarise the results of that particular study, we found that datafication is very often seen as something that changes the interface, okay?" (I4, Pos. 9).

In the above quote, I4 refers to the 'interfaces' of interactions between citizens entitled to certain public services with the programme of these services' provision, explaining how technologies at large and data infrastructures in particular become part of this 'interface' where previously only different people were interacting with each other. Continuing this argument and in the spirit of sociological structuralism, I19 elaborates on their view of data infrastructures as enabling dynamic datafication processes. I14, in turn, underscores the distributed character of data infrastructures, as these are usually understood as big scale technologies comprised of a manifold of actors, both collective ones (e.g. public administrations, organisations, NGOs who own, process, overwatch, and extract data from these infrastructures), non-human actors essential for connecting these different actors to one another and to individuals (e.g. citizens) who utilise these data infrastructures.

Researchers' access to the data infrastructures and the manifold of involved actors, however, is often limited (physically, politically, temporally). Thus, some of the interviewed scholars turned to policy research to trace the dynamics of data infrastructures through publicly available materials (websites, documents, recording of events, etc., e.g. I4, I15). While this kind of research design is primarily central to the methods assemblage of reconstructing datafied regimes, discussed in the next section 6.4 of this chapter, it also allows to trace historical dynamics, predecessors, and future imaginaries of data infrastructures. In addition, as I15 discusses, policy and document research allow exploring dynamic imaginaries of data infrastructures across various stakeholder groups. In I15s work beyond the European context, media reports were used as a way of cross-checking information about data infrastructures publicly available on the government's websites and documentation provided by the authoritarian government in research situations in which I15 had no empirical access to the data infrastructures themselves.

> "And so, I'm going first always to the ministry […] website and documents and trying to dig out that kind of information. Increasingly also like they publish news releases and so I'm trying to also then follow those up. Often, recently I'm also watching conferences, like you can find different ministry organized conferences or interviews. So, also, kind of following them up. Even like digging out presentations that civil servants have held on the topic that interests me. But then also kind of, I think media is really interesting if you have, especially like different kinds of newspapers. Like government friendly, more critical, more like -, just like very popular, you know, like easy going newspaper. So, I'm just trying also to see -, sort of crosscheck maybe. If the government paper is saying this, is it also still like what the critical paper is saying? So that also, for instance, if I'm trying to more like establish the factual information, that I'm not completely kind of wrong, you know, that I'm maybe following false information that the government wanted to spread, whereas the critical newspaper has actually criticised something." (I15, Pos. 9).

Few other experts, however, traced dynamics of data infrastructures computationally, identifying, visualising, and exploring the links between multiple human and non-human actors (devices such as smartphones, apps and software processing e.g. use data, data centres storing these data, and big tech corporations, public bodies, and intermediaries who own, sell, and use these data for their own goals, e.g. I14, I16). Other methods, such as data journeys have been mentioned by some of the interviewed experts (e.g. I16, I18, I26) as an inspiring methodological approach and further possibility to trace dynamics of data infrastructures by following the movements of data across these. I26 summarises that as follows:

> "You could make that a method, so by following the data around these journeys and stopping off at different places and spending some time there and looking at what was going on and thinking about how these different people that are embedded in these different places and all that values and things like that helping to push the data on to the next place, or maybe stop the data going on to the next place" (I26, pos. 5).

As the quote illustrates, some of methods applied by scholars tracing dynamics of data infrastructures put the notion of dynamics in the foreground and specifically acknowledge how digital data move across data infrastructures between various actors. These actors, in turn, work 'behind' data infrastructures in order to keep digital data moving. Data infrastructures, then, can be understood as spaces where data travel from one actor and information systems to another.

One line of research on tracing the dynamics of data infrastructures, therefore, specifically attends to tracing the movement of data across infrastructures (I14, I17, I19, I26). While most experts acknowledge that data are "moving around all over the place" (I26, pos. 5) across data infrastructures, connecting local stakeholders and physical, material devices and sensors required to capture data all the way up to the global markets and corporations making predictions based on the processing of these data. Thus, tracing dynamics of data infrastructures based on following data movements provides an illustrative example of how data infrastructures connect a big number of different actors, including individual, collective, and non-human actors together. The focus of tracing data movement is, then, not solely on producing a detailed account of all passage points, but rather on the value-laden work practices that allow data to move or not; on the socio-cultural, socio-political, and socio-economic factors that push data forward in a certain way or prohibit data to move forward; on the practices and actors that change data and make them come to matter in different ways at different points of passage within the data infrastructures (I26). Several experts (e.g. I14, I30) identified data centres as points of the data infrastructures which digital data moving globally pass through. For I14 and I30, exploring data centres allows some insight into relations between various global actors, for example social media sites and stakeholders processing, selling, and reselling social media data. Others turn to the material aspects of the data movement across the infrastructures, such as bandwidth and sensor capacities as well as the architectures of data infrastructures (I17). Another way to capture data movements is to identify actors between which data move, and their spatial proximity (I19, I22). The spatial boundaries between actors (e.g. organisations) may play some role there even if data movements do not completely adhere to the spatial distribution of actors alongside data infrastructures. As I19 exemplifies, however, sometimes the proximity of certain actors impacts data movements, as e.g. informal conversations and personal connections may e.g. speed up or change original data movements. Attending to actors connected through data infrastructures also helps to examine the origins of datasets that are being moved and used across data infrastructures (I22). Attending to the origins of datasets and data movement also means attending to the politics of data movement (I26). For example, researchers ask who is controlling data movement, which passage points do data go through while moving across data infrastructures, what values are "helping to push the data on to the next place, or maybe stop the data going on to the next place" (I26, pos. 5). Another way to explore data infrastructures is to attend to the lacking data movement in the situations when this movement would be expected by the stakeholders at the site of practice and the lack of movement disrupts practices possibly leading to dire consequences (I29). Although the quoted experts underscore the importance of tracing data movement, others (e.g. I18) warn against a solely empirical focus on following data movements and the lack of conceptualisation required to understand the complexities and the non-linearity of data movements. Data movement and the work required from actors for data to move, therefore, hold data infrastructures together and connect their various actors. As I26 elaborates, sometimes these data movements are interrupted and frictions occur, for example when the movement stops, pauses, or is being redirected (e.g. Edwards 2012, Bates 2019). By studying these movements and frictions scholars tracing dynamics of data infrastructures examine how these movements and frictions create relations, including and excluding various actors, including and excluding data or goods moving across infrastructures. As these inclusions and exclusions are enacted in accordance with political, economic, cultural, and ideological specifics in which data infrastructures operate, datafication researchers tracing dynamics of data infrastructures require deep tacit, practical, and professional knowledge about the sites of practice they investigate

empirically. The combination of these knowledges and dynamic units of analysis allows scholars enacting the methods assemblage discussed in this section to understand better the implications of datafication processes enacted through data infrastructures.

6.3.3 Dynamics of data infrastructures and other methods assemblages

Some of the scholars whose research can be understood in terms of methods assemblages of tracing data infrastructures report that datafication processes have not been their initial research focus (I13, I15, I17, I27). Some of studies tracing dynamics of data infrastructures initially begin as research projects aimed at reconstructing datafied regimes inscribed in certain software and apps, as discussed in the next section 6.4 of this chapter. In the course of these projects, however, researchers switched their attention to the moving of data beyond one information system and across data infrastructures. As I14 described it, their project was aiming to "kind of open up the hood and look at that infrastructure that is sort of moving continuously and reshaping and reforming itself" (pos. 13). This reshaping and reforming takes place as different kinds of actors are situated in the performative relations with data infrastructures, while software, apps, and algorithms are among the many non-human actors. To be able to disentangle these relations, researchers need to look "beyond the hype, beyond the popular discourses, beyond what the tech companies are trying to tell you about data." (I21, pos. 27) For example, I14 discusses how examining 'strategic importance' of data and the decision-making about how data are distributed across data infrastructures helps to understand which of the relevant actors hold agency and power. Some scholars, therefore, attend methodologically to these power dynamics in different ways. I8 also asserts that in contrast to ethnographic observations, interviews allow to understand the power dynamics better. In contrast to that is the point made by I30, working ethnographically, who argues that ethnography allows understanding infrastructural labour of actors who are usually not included in research, such as professionals who make everyday decisions about the maintenance of data infrastructures and are involved in the everyday activities of setting data on the move.

Similar to the methods assemblage of exploring encounters with data representations, scholars tracing dynamics of data infrastructures are also interested in the ways individuals and collectives utilising data infrastructures encounter these, for example by interacting with certain technological 'interfaces' (I4). Their analytical attention, however, is directed at the various implications of data infrastructures for the stakeholders, while these implications become visible and assessable as inclusions and exclusions made through data being moved across data infrastructures. Through a long-term commitment to tracing dynamics of data infrastructures, researchers acquire tacit knowledge and access to the spaces of inclusions and exclusions, required to situate datafication processes enacted through data infrastructures empirically. With this section, I aimed to illustrate how datafication researchers tracing dynamics of data infrastructures are analysing implications of these infrastructures through analytical attention to data movement and the work required from various to keep these data on the move. My purpose here was to flesh out what is dynamic about data infrastructures, and how these dynamics can be traced methodologically in order to produce situated knowledges about data moving across these infrastructures and datafication processes this movement enables. Data movement and work 'behind' the infrastructures seem to be crucial here as through this movement various actors and elements of data infrastructures relate to each other and exclusions/inclusions of these actors and elements are enacted. For understanding datafication processes, then, the core questions are about who does the work of moving data, how are they related to actors utilising data infrastructures, and what kinds of relations between various actors this work aims to establish? Before I continue discussing how the heuristic of methods assemblages developed in this thesis furthers our understanding of datafication processes, I elaborate on the third methods assemblage of reconstructing datafied regimes also investigating work 'behind' data infrastructures and software.

6.4 Reconstructing datafied regimes

In the previous section I elaborated on the methods assemblage of tracing dynamics of data infrastructures allowing to explore social, economic, and political implications of datafication processes. The methods assemblage of reconstructing datafied regimes, in contrast, sheds light on the negotiations of data representations by actors in positions of enabling datafication processes, for example software and technology providers or policy-makers. The first subsection elaborates on the practices of such actors. The next subsection elucidates how enacting this methods assemblage datafied regimes can be understood, followed by a detailed discussion of how datafication scholars view negotiations of data representations in datafied regimes. The two final subsections introduce research procedures and techniques applied for reconstructing datafied regimes and delineate this methods assemblage from overs discussed in my thesis.

6.4.1 Studying data production

The method assemblage of reconstructing datafied regimes addresses practices and work that may be referred to as data production (e.g. decision-making about design, development, maintenance of information systems). Depending on the studied empirical site of practice, some interviewed scholars identify documentation (e.g. policy documents, software specifications) as their central objects of study (e.g. I2, I10, I15). While software documentation here can be considered as a more immediate source for reconstructing datafied regimes, expert interviews suggest that in many cases policy documents can be seen as regulatory frameworks guiding the application and usage of algorithms and digital data. Others attend to software and algorithms themselves, reverse-engineering (see e.g. Bucher, 2016), auditing (e.g. Sandvig et al., 2014; Shen et al., 2021), or 'walking through' (see e.g. Light et al., 2018) these in order to reconstruct the datafied regimes inscribed in these (e.g. I21, I28). The multiplicity of terms used by the experts to describe different kinds of information systems, ranging from tools, to software, to systems, to platforms, to algorithms indicates 1) different disciplinary and theoretical backgrounds of the experts and 2) the lack and impossibility of a unified definition.

As I have shown in chapter 5, though, document and policy research differ significantly from other techniques like a walkthrough, also in regard to the matters of concerns these techniques help scholars raise. What is common, however, and allows to bring both together under the umbrella of the methods assemblage of reconstructing datafied regimes, is the shared analytical attention to different practices of enabling and producing datafication processes, to the negotiations about how meaning is attributed to data representations. These negotiations take place on different levels, both within commercial organisations negotiating on the development and design of their products and within other actors such as political ones. Common to all datafication researchers enacting methods assemblage of reconstructing datafied regimes is an acknowledgement of technologies (e.g. algorithms) or documents such as policies as socially constructed and including certain assumptions about realities of the people making decisions about these technologies or policies. This argument is summarised by I28 in the following with focus on the concept of "social reproduction". Drawing on other interviews (e.g. I17, I19) the concept of "reproduction" can be questioned in favour of other concepts addressing how within socio-technical systems society and technology are constantly reconfiguring each other. Nevertheless, the following quote is quite illustrative for the argument on the construction of algorithms.

> "And for example, I don't know if you know sociologist Bourdieu […]. So, he has the concept of cultural reproduction that the kids grow up and they learn to reproduce socio-political structures, which they learn at school and the family consciously or unconsciously. […] So, the algorithm is a subject that was trained then shaped by the decisions of the company of the engineers and the data it was fed, and then leads to some socio-political reproduction." (I28, pos. 11).

In a similar way as technology developers negotiate the categories required to represent the realities in these, so policy documents provide an insight in the accomplished negotiations of the same question by policy-makers. For example, for I2 policy documents were among the main resources for their empirical research as these "express ideas that are outside kind of the borders of where the document was written" (pos. 17). I19 explains their methodological approach to reconstructing datafied regimes as centring an analysis around a certain software or 'tool' that is used by some actors to analyse, produce, and negotiate data representations. They also elaborate on the analytical discrimination between data use and data production, noticing that practically these two kinds of practices are often not separable.

> "And I think, use is something that happens already when you operate a tool […], when I explore the design and when I am located in the context of use, or I anticipate this context, or use it [a digital tool] for myself or to show something someone else. So, this kind of use is already within production and they cannot really be separated. And that's what I am interested in understanding, what does it mean. And if I take this tool, […] as a pivotal element and place it in the middle [of analysis] and then try to understand different practices around it." (I19, pos. 9).

6.4.2 Understanding datafied regimes

Decision-making of designers or policy-makers and their assumptions about the world inscribed in software or policy documents are the central unit of analysis in projects reconstructing datafied regimes. To situate these decision-making practices empirically, scholars applying this methods assemblage question "to what extent do these assumptions are born out in practice" (I13, pos. 17). For example, while it might be one group of designers creating a certain app, they use developer tools provided to them by other companies, interested in data extraction (e.g. I14). Thus, decision-making here is distributed across multiple organisations and individuals, as well as technologies that create affordances for data production. While practices of data production described here are somewhat similar to the development and maintenance of data infrastructures described in the previous section 6.3 of this chapter, researchers enacting methods assemblage of reconstructing datafied regimes are specifically interested in the decision-making practices required and enacted by various stakeholders as they 'produce' data. Like the methods assemblage of exploring encounters with data representations, scholars reconstructing datafied regimes attend analytically to the translations in which datafication processes are being enacted. Experts reconstructing datafied regimes, however, investigate how actors in positions to enable datafication processes (such as software designers or policy-makers) make these translations and inscribe these in products of their work (e.g. technologies or policies).

When such assumptions are inscribed in technologies (e.g. algorithms) or policies, they become normalised as datafication processes are being enacted through these software or policies. For example, I2 provides an example from education research in which certain assumptions about education quality assessment are normalised through education policies: "the discourses of quality, they are permitted by data and datafication nowadays" (I2, pos. 11). Education policies examined by I2 determine practices of other actors, for example educators, who need to adopt policies in their professional routines. According to I2, for policy-makers datafication processes, then, can be understood as a solution for better policy-making as it provides interoperable, quantifiable data representations the meanings of which have been negotiated according to the "policy scripts" (I2, pos. 15). These policy scripts define how the realities of education (in the context of I2's research) should be represented in digital data. In short, I2 views policy-making as circulating and acquiring information. Similar to I2, I19, also applying methods assemblage of reconstructing datafied regimes to study educational policies, elaborates on the example of education quality assessments how these policies are normalised as they cannot be changed through individual will.

> "So this is, this is so to say basic sociology. So, there are structures, these structures are supra-individual and they take effects and do something and are dynamic and have ordering forces which are not managed purposefully by other actors, but develop certain power. […] And the more I do research, the more I see this. The actors can be so reflective and try to curb some things. I agree here with Zuboff, if data are there, they will be used. If data are embedded in a certain structure, something will happen with them. Then the dynamics of centralisation, dynamics of expansion, and of accountability are under way that I cannot prevent." (I19, pos. 15).

Particularly in the example of education quality assessment studied by I2, I5, and I19, policies enacting datafication processes reinforce accountability and transparency as "modern principles of governance" (I2, pos. 19). I29 offers three different views to understand datafication and policy. First, an "old Anglo school lens of modes of power and understanding how datafication is used overtly to make people do things that they otherwise wouldn't do. So that's in a regulatory government sense". Second, "[i]n covert ways where they basically inform the agenda and ways of knowing the world" and third, a political view on "how information and knowledge are related to relationships of power" (all quotes pos. 19). I5 furthers this argument pointing out how whole educational organisations and not only individual educators become subject to these policies, powers, and 'dynamics'. Others, for example I16 and I25, address these 'dynamics' as 'logics' that enact data representations of people, things, and practices. So, I16 asks, "what kind of logic dictates in the systems and how our population is understood, our people, individuals and social issues understood" (pos. 13) by public administration in a European country.

Assumptions about datafication inscribed in algorithms or policy documents, then, become 'ordering forces' for other actors' practices, as mentioned by I19. Scholars enacting methods assemblage discussed in this chapter, then, investigate products of decision-making processes such as software documentation or policy documents, or sometimes also practices of other, affected stakeholders in order to reconstruct these 'ordering forces'—datafied regimes. The variety of definitions ranging from dynamics and scripts to logics illustrate the heterogeneity of theoretical and epistemological approaches used by the experts enacting this methods assemblage. While these different terms bring forward different aspects of what I call here 'datafied regimes', they share a common perspective on datafication processes as a political endeavour negotiated and articulated in policy and governance strategies, be these educational policy documents (e.g. I2, I15, I19) or platform use policies (e.g. I14, I21, I25). I14 also touches upon the question of how 'normality' is negotiated in and defined through datafied regimes or what happens if the understanding of 'normality' changes. In I14's example, what is 'normal' does not concern people or their practices, but rather different kinds of data tracked and extracted by social media providers.

> "But we also kind of realised in kind of looking through kind of technical literatures that -. And I think a lot of work needs to be done here, but we at least sort of bring attention to, like something like [certain datapoint], which is in like 99% of apps is […] normal […]. This used to be [categorised as a] dangerous [datapoint]. So, it was changed to normal. And based on everything that we looked at, I -, my hypothesis is, it was changed." (I14, Pos. 11).

Here, what is understood as literally normal kinds of data extracted from social media and app users has been allegedly changed in the course of several years, according to I14, also changing, for example, whether or not users are being informed about these kinds of data about their usage being collected. I25, in turn, elaborates on one of their research projects with software providers and data scientists, also reporting how certain understandings of data and datasets change over time, reshaping what is understood to be 'good' or of 'high quality'.

> "In this text they explain… and I… it is about language. They explain that they [(software providers)] have these data. And they have these data and maximum 40% of these data need to be there to speak about a dataset of high quality. And somehow two paragraphs below this is a dataset of high quality. And this transformation is so natural. So natural in the logics

> of data science because other way you cannot do anything [with the dataset]. But the way it is translated to a completely different understanding about what these data can represent, this is […] So about developers… It is what fascinates me about such developers' perspective, what is natural for them." (I25, pos. 23).

In another example from education research, I5 elaborates on categories and metrics used for education quality assessment reconfiguring the whole learning process: putting students in different rooms literally and in different boxes metaphorically, respective to the categories attributed to these students. Similarly, I8 reports their experiences and frustrations about "the kind of pointlessness of it, you know, that this is just making numbers and moving them around." (pos. 25) In sum, scholars enacting methods assemblages of reconstructing datafied regimes turn to the question of how datafied regimes produce 'normality'. Other scholars take that as a starting point and make calls for new regulations that would restrict the 'dynamics' of datafication processes enacted through datafied regimes (e.g. I6, I21, I23). At the same time, some of the interviewed experts such as I6 argue that currently not even governments can be held accountable for their datafied regimes they negotiated because decision-making processes making these regimes possible are not transparent and not known for the publics. Similarly, I23 argues that transparency is required for researchers and civic society cannot understand what governments and tech corporations do with the data they accumulate. I11, in contrast, draws attention to another kind of policies, examining empirically self-regulation policies concerning datafication processes in certain professional community (e.g. ethics codes) and the ways datafication processes are being enacted in these.

6.4.3 Negotiating data representations

For most of the interviewed experts who enact at least some elements of the methods assemblages for reconstructing datafied regimes, central to their re-situated conceptualisation of datafication is a critical reflection on how social phenomena (e.g. learning, citizenship, but also e.g. emotions) are defined through datafied regimes. For example, I2, I5, I8, I10, and I19 elaborate on policies and governance enacting certain datafied regimes, while at the same time being themselves dependent on datafication processes in various public societal domains such as administration, health, or education. For example, I4 reflects on "the policy change enacted with datafication" (pos. 9), while I8 notices that "a lot of datafication arises from policy, certainly in [Europe] anyway" (pos. 23). Enacting methods assemblages for reconstructing datafied regimes, therefore, datafication scholars explicitly investigate the recursive and political relations between datafication processes and regimes that enable and enact these. As I27 summarises it, "technology enshrines, it literally physicalises all of those decisions, the ethical ones, the scientific ones, and all the other decisions" (pos. 10).

This chapter illustrates how data representations negotiated by various actors bring forth expectations of acting upon them. In a similar way, some of the interviewed experts argue, datafied regimes do not only define realities according to previously negotiated, politically- and value-laden representations of data, but also create affordances for certain actions. For example, certain digital data representations are being enacted as 'objective' and 'truthful' (e.g. I2, I15). I2 elaborates on that on an example of their research on education policy and policy-makers' decision-making practices that rely on certain data representations. I5 provides another example, arguing that datafied regimes create certain affordances for educators' actions and these actions are sometimes directed at the data representations rather than students' learning.

> "And then, I see a very important role of numbers, and numerical data, and objectification in kind of like creating these new actors, and these new roles, and like this kind of political arena. Kind of like numbers, like to be more neutral actors and non-subjective, because there is lots of criticism for subjectivity when making policies or in policy-making. And that's why numbers have been used to make the decisions more objective. And then numbers sometimes, are indicators or rankings. They become actors themselves, like without a

> corresponding human being behind or a corresponding organization or something like that." (I2, pos. 17).

> "[T]eachers are heavily influenced by the way in which [educational organisations] have gone about making things of what it is that they have achieved in the past decade. And it is a lot of focus on different sorts of graphs and charts, and metrics and measures of students learning as such. And it can almost be the case that these graphs and charts kind of dominate what's actually occurring in relation to the students learning. So, you know, what you end up with is more focused on something like these graphs, as indicators or approximations of student learning, then you do upon the actual substance the student work. So, I think that's risk. And I think that's a risk at the level of systems, of [education] systems around the world." (I5, pos. 29).

Taking up on the argument that enactment of datafication processes through datafied regimes in policies creates certain affordances for the actors subjected to these policies, I19 reflects on the role of policy-makers. In their research, I19 notices reflexivity of policy-makers regarding their power positions and actionability in negotiating and defining datafied regimes.

> "There was barely anyone with whom I've spoken in the context of governance throughout my research, who wouldn't have been reflexive about what these things can do, what are the dangers, the opportunities, what is their own role there. Even when it wasn't completely analytical or systematic, it was almost never anyone who was a member of such a structure or worked with such a system und haven't thought about it at all. And particularly when it is about designing something, there are so many processes, structures, actors, revision loops. […] rather they aim to do a good job within their domain of authority. And there happens a lot and it casts certain light on the critique. […] because this critique is about the dynamics of the process, about the challenges of datafication and not about people being stupid or not being reflective about what they do." (I19, pos. 15).

As I19 further argues, however, it is difficult for the involved actors to situate data that is used for governance and policy-making empirically. At the same time, as I30 reports from their work with policy-makers and software designers, sometimes their decisions about negotiating data representations are made to provide only particular options of acting upon published data, for example to "make sure that journalists will not take pictures that will make a headline" (I30, pos. 7). However, I30 also warns against being exclusively critical towards the designers' practices and elaborate on the "contingencies, all the choices you have to make when you develop a [data representation]" (pos. 3).

Who gets to decide and negotiate data representations, then, is a political question relevant for reconstructing datafied regimes. As previous examples provided in this section demonstrated, most of the interviewed experts reconstructing datafied regimes either conduct policy research or explore professional practices of policy-makers and software designers, or both, sometimes complementing these with examinations of software interfaces. That overview over scholars' data collection and analysis techniques illustrates what kinds of actors are considered as holding power and access to the negotiations of datafied regimes. Policy-makers, software developers, and intermediaries are understood to have various ways of exerting their agency and power in negotiating and defining datafied regimes (I21). Particularly tech corporations or governments accumulating data through extraction and negotiating representations of these are seen as holding positions of power. It is important to acknowledge that these different actors often pursue their own agendas and goals. For example, I15 reflects on how "suddenly, [data] become meaningful in new ways and new contexts" (pos. 3) as they require to be acted upon, in their research also in the domain of education governance and decision-making. Thus, in their studies reconstructing datafied regimes of education governance, I15 is particularly interested in the arguments that political actors use, for example in policy documents, to justify datafied regimes they negotiated. I17

provides an empirical example of the relations between datafication processes and politics. This example illustrates negotiations about data representations for decision-making in a public domain and questions how various actors perceive datafied regimes and what is the role of data representations for decision-making processes in the first place. In I17's research, public actors entered conflicts with each other during negotiations over data representations in order to acquire or maintain their range of influence.

> "I have participated in meetings in which they yelled at each other and were completely split. How to go about these data? Who should benefit from these et cetera? But often also, sometimes really childishly, so insane jealousies developed there. Who has access to which data? What does it mean to lose the range of influence? Often it went as far as whole databases should include the names of certain actor groups. Yeah, so, completely detached from the actual goal and content" (I17, pos. 7).

The above quote by I17 illuminates how datafied regimes are enacted besides strategic negotiations of data representations, for example by allowing or closing access to certain data representations. I4 also makes a similar argument discussing how datafied regimes structure who gets access to other datafied practices such as, for example, public services. I23 reflects on pitfalls and challenges of private companies regulating access to negotiations about datafied regimes and to the data generated within these regimes, particularly in the public health domain.

> "I mean, I don't want to know the formula, but I want to know the criteria, at least, that are used to profile users, to take decisions, to do a lot of things. And of course, one of the- I mean, I don't know the specific research on that, but one of the main areas where datafication is important of course is health. […] we could do a lot of things with data with regard to health, but of course we cannot leave all this data just to private companies. So at least in Europe, we would need regulation, but in order to have all this stuff, you need politicians and decision makers that understand what's at stake. And maybe citizens don't, but I'm not sure, at least in my country, politicians do. So, we have a great resource there, but we are not using it. (I23, Pos. 17).

In their research I27 experienced problematic ways of operating with data in the educational domain and how these were negotiated by involved software providers.

> "Yet even then, when I was actually speaking with […] these learning environments, and some of the research questions on the back of it with my colleagues, you again have to make some kind of a philosophical decision on whether you are comfortable with seeing all learning as consciously, explicitly, articulable, decomposable knowledge. And if so, if you're also comfortable with designing systems, and using people, human learners, as ways to investigate, and to evolve that knowledge mapping." (I27, Pos. 8)

In addition to previously discussed challenges in regard to datafication processes enacted through datafied regimes inscribed in software or policies, this quote also draws attention to the ethical questions of using available data for negotiating further, future data representations and datafied regimes. In another example, I21 was interested in identifying datafied regimes of social media platforms.

> "I want to look what is the shaping role that platforms play generally in users lives, sometimes in economies and broader social structures. And so, I start from there and because of the way platforms work, because of the way they make their money, datafication is a big part. So, looking and identifying how are they rendering everyday life into metrics and numbers and data that can then be repackaged and sold onward and so yeah it is something that affects users. It's something that affects organisations." (I21, Pos. 3)

As the above examples suggest, methods assemblages of reconstructing datafied regimes primarily address various collective actors such as public administrations, software provider companies, but also organisations subjected to various datafied regimes. This attention to collective practices distinguishes this methods assemblage from the other ones, discussed in previous sections of this

chapter. As the calls for regulation, mentioned by several experts in the examples throughout this section demonstrate, collectivity of actors negotiating data representations and, with them, datafied regimes, also poses legal and ethical challenges, for example in relation to the questions of accountability. I continue discussing the role of collectivity in studying datafication processes in chapter 7 of my thesis.

6.4.4 How are datafied regimes reconstructed?

Overall, empirical examples discussed so far elucidate how scholars enacting methods assemblage for reconstructing datafied regimes study practices of data production: from planning and decision-making, to development, implementation, and maintenance of information systems. Within this methods assemblage, a number of techniques are used that can be summarised as software analyses (e.g. walkthrough, algorithmic audit, reverse engineering, eg. I19, I21), interviews with technology providers and other decision-makers (e.g. I25, I27) or actors subject to the datafied regimes (e.g. I4, I17), and document or policy research (e.g. I10, I15). For example, I25 views interviews with software developers as exploration of the inscriptions they make into their products, and therefore as an exploration of software. At the same time, this view is contested in the literature (e.g. Hasinoff & Bivens, 2021). So, according to I30, interviews with software designers allow to situate findings from other methodological approaches (software analysis, document analysis) as they provide researchers with more information than what is possible to learn from official, written, publicly available accounts. For example, interviews allow to learn about technology designers' negotiations on how to deal with certain problems, conflicts, and missing data. They explain that "sort of historical projects that never came into being because of fights between the ministries, they are forgotten. They have been written out of history" (pos. 17). I30, however, also warns against being too critical towards the study participants. Instead, the focus should be on

> "understanding how the people who work with it, how they perceive it, what they think about it, what they think its problems and potentials are rather than sort of, you know, using them as puppets to confirm my own critical position" (I30, pos. 15).

According to I30, datafication scholars should engage in mapping the complexities in which different kinds of actors (both human and non-human ones) move and make decisions about datafied regimes. At the same time, I30 highlights a challenge of access to research materials in ethnographic studies with software providers and policy-makers: "It's also because when you're doing ethnography, you have access to what they invite you to have access to" (pos. 3). I6 makes a similar comment, also highlighting that even when access to relevant actors is acquired, researchers might not be able to acquire information they need: "sometimes it can be difficult to navigate, because sometimes they don't want to talk about certain things, because it's just scandal or they are trying to move on to something" (pos. 9).

The issue of accessibility of research materials and units of analysis, therefore, is one of the central methodological challenges for the methods assemblage for reconstructing datafied regimes. Researchers reconstructing datafied regimes by examining technology and algorithms in most cases face constraints such as lacking access to the empirical data (e.g. technical documentation or access to meetings in which certain aspects of these technologies are being negotiated) required for the analysis, lacking access to relevant stakeholders, or documentation. Algorithms used in different kinds of information systems are often proprietary, undisclosed to the researchers.

> "So, especially when studying how the algorithms transform politics, the fact that they're either proprietary or undisclosed means a lot of times that the research community performs for years studies that the tech companies know the results and they don't share them. Or they are not able to investigate really important to see the issues because they don't have access to the data. This is my biggest concern with datafication." (I28, pos. 17).

Moreover, decision-making processes are usually documented to a varying degree of completeness. Some of the interviewed experts, therefore, turn to studying experiences of people affected by

datafied regimes as a helpful addition to the methods assemblage, also combining elements of different methods assemblages such as exploring encounters with data representations.

Some of the experts whose research projects I identified as reconstructing datafied regimes, therefore, apply policy research and document analysis techniques. I15 elaborates that "working with documents and media materials is my way to kind of substitute the difficulties of field access" (pos. 9). For I15, primary sources of documents are political actors (such as ministries) who publish official documents. In addition to these, I15 increasingly also examines other kinds of publicly available documents such as public talks, recordings of events, interviews in press and media coverage following up release of official statements. Further, I15, working in policy research in a non-European country, also conducts analysis of media coverage on the topics related to the datafied regimes identified in the policy documents. According to I15, such combination of various research materials allows to compensate for the limited access to the political actors themselves (e.g. for conducting interviews with them). I25 conducts discourse analysis of press and media texts, however they focus on the perceptions of the datafied regimes, particularly attending to the language used in the studied documents, for example exploring what is presented as 'problematisations' or 'solutions' in these texts. Attending either to problems or to solutions proposed in texts about what they address as software allows I25 to identify core values central for the negotiations of datafied regimes. I25 also elaborates on advantages of interviewing actors directly involved in negotiating data representations performed in datafied regimes: I25 relies on their "ethnographic sensitivity" (pos. 7) to topics emerging in loosely structured interviews with software designers. In their research project, I16 who also works outside Europe, could complement their policy 'desk research' with interviews with developers of software for public authorities. They also notice that the multiplicity of terms, concepts, and their meanings used in documents created by the study participants pose a challenge for researchers who might use the same terms or concepts differently.

Other scholars applying methods assemblage for reconstructing datafied regimes conduct various kinds of software analysis. For example, I19 elaborates on how software analysis (with particular focus on dashboards) complemented by interviews with software designers allows to understand the 'logics' of the latter and the datafied regimes inscribed in these. I19 states that a software analysis allows "navigating the platform with the pace that research allows." (pos. 11) In addition, for I19 technical documentation can be another helpful source for software analysis, especially when interview partners cannot be accessed. I21, who conducts another kind of software analysis, notices how conducting these for apps on mobile devices requires taking into account various aspects such as current system updates and app updates. In addition, personalised apps (e.g. social media) also complicate research: for example, when starting new research topic lacking use history hinders researchers in their analysis of social media algorithms (I21). Moreover, especially conducting research with mobile devices, sufficient funding is required to accommodate various expenses, for example for devices. Another way to study decision-making when proprietary algorithms and information systems are not accessible and assessable for researchers is to recreate some of their products, for example to study how different kinds of algorithmic trace data lead to certain outcomes of machine learning-based automated systems (I20). I24 points to a limitation of such kinds of software analyses, elaborating on how, for example, analysing app interfaces only serves to answer research questions about these particular apps but not the datafied regimes according to which these apps are displayed in the Play- or Appstore: "Obviously you cannot have that [ecological perspective] by taking screenshots of the app" (pos. 3). Further, I24 also underscores the value of disciplinary perspectives for the software analysis, highlighting how apps for health, education, some everyday activities cannot be compared to each other analytically directly, but rather require working out specificities of each societal domain based on academic discourses respective research fields.

Some other interviewed researchers use computational techniques of data collection and/or analysis as a part of the methods assemblage for reconstructing datafied regimes, also reflecting on the limitations of such techniques. For example, I16 reminds of the trends in academic inquiry: "popular computational methods around, was the topic analysis and social network analysis and things like that" (pos. 27). According to I16, when a computational technique becomes 'popular' it is quite important that theories in which the development of such techniques was grounded are taken into account and reflected by empirical scholars applying these techniques. Similarly, an overall understanding of what comprises 'software' is crucial for those who aim to study datafied regimes enacted in it.

> "[Y]ou require technical knowledge, I think, like what is a platform you can talk about that if you don't at least have a bit of technical knowledge about what it is like, knowing that it consists of GUIs and APIs and whatever like that, you have an idea about what it is that if you do enough study or whatever, to come back to that example that you at least know that the Play store is closely connected to that and has some role to play in what is this app is on the forums and that sort of stuff" (I24, Pos. 21).

For an "average lay user" (I21, pos. 31) algorithms are often difficult to understand. Some of the researchers applying methods assemblages for reconstructing datafied regimes, therefore, also aim to enable their study participants and increase their algorithmic and data literacy through research. For example, in one project conducted by I23, they and their colleagues explained their study participants "what we knew about at least [the app's] algorithms" (pos. 19). In this research project, some of the study participants already tried to 'reverse engineer' algorithms of the apps they used themselves by trying out different app use strategies and comparing the results of algorithmic recommendations. These experiences were incorporated in the research processes as I23 explained in the interview. While the goal to empower study participants is common to the methods assemblage for exploring encounters with data representations and its participatory approaches to research, the example provided by I23 rather indicates how academic inquiry becomes more situated and partial, taking positions and stands vis a vis development of datafication processes, strongly enabled by data extraction, exploitation, and systemic inequalities. Such positions and perspectives of researchers also become an inherent part of research even when datafication processes initially are not the main focus of study, as was the case for e.g. I2, I13, I15, I17, I21, I25.

6.4.5 Datafied regimes and other methods assemblages

The examples of research projects discussed in this section as enacting the methods assemblage for reconstructing datafied regimes, several scholars enact elements of this methods assemblage together with elements of other methods assemblages, discussed in previous sections of this chapter. For example, I21, has an infrastructural view on social media platforms and explores empirically datafied regimes inscribed in these platforms. They explain that "platforms take some responsibility in conveying that [information] rather than just users circulating it" (I21, pos. 3), while this responsibility comes from the monetisation of social media data travelling across data infrastructures. Here, the experts' main research interest is in understanding datafied regimes enacted through social media platforms, while these platforms operate on vast data infrastructures, ultimately, required to monetise data.

> "I want to look what is the shaping role that platforms play generally in users lives, sometimes in economies and broader social structures. And so, I start from there and because of the way platforms work, because of the way they make their money, datafication is a big part" (I21, pos. 3).

Similar to how scholars applying the methods assemblage of tracing dynamics of data infrastructures, I21 argues that without being able to know how data move across these infrastructures, democratic values cannot be enacted in datafied regimes of the platforms. Some other examples of research projects and academic inquiries in which elements of multiple methods

assemblages are applied provide interviews with I4, I4, or I17, whose—primarily ethnographic—work.

Overall, methods assemblage of reconstructing datafied regimes resembles the previously discussed methods assemblage of tracing dynamics of data infrastructures due to the analytical attention to the work practices of collective actors—policy-makers or technology providers. Specific to reconstructing datafied regimes is the interrogation of how such datafied regimes come to be rather than their implications on stakeholders subjected to these. Due to the limited access to the empirical sites of practice at which data representations included in the datafied regimes are negotiated, some scholars explore stakeholders subjected to these regimes, such as e.g. educators in research of datafied regimes in education governance. Most common, however, is an analysis of either direct situations and experiences, perceptions of negotiating data representations or of documentation of these negotiations. Despite such different research techniques, datafication scholars enacting methods assemblage of reconstructing datafied regimes explore how in these regimes digital data are attributed to what is understood to be 'normal' or not. In the concluding section of this chapter, I elaborate in more detail on the core aspects of the methods assemblages presented in this chapter, the differences, and relations between them.

6.5 Assembling methodological approaches to data studies

In this chapter I analysed expert interviews with datafication researchers about their empirical studies and based on my analysis constructed three distinct methods assemblages currently enacted to understand and create re-situated concepts about datafication processes. The discussed methods assemblages are applied for 1) exploring encounters with data representations, 2) for tracing dynamics of data infrastructures, and 3) for reconstructing datafied regimes. These methods assemblages are, themselves, related to broader research politics in social sciences. The three constructed methods assemblages, while being distinctive from one another, overlap and form a continuum of methodological approaches within data studies. They can be conceived of as synthetic conceptualisations of the methodological approaches identified in the current body of work on datafication. These three assemblages by no means form a comprehensive list of all possible methodological approaches, rather they illustrate the results of my analysis based on a limited sample. The three constructed methods assemblages are grounded in the samples of academic literature and expert interviews I examined in my thesis. With further research on methodologies of data studies that listing of assemblages might include other, related topics and with the development of datafication scholarship over time new assemblages might be added and existing assemblages reconstructed.

Each of the methods assemblages discussed in this chapter has their own distinctive characteristics. The methods assemblages overlap and relate to each other as well as to empirical research in social sciences in general. So, interdisciplinarity of research teams and disciplinary boundaries of certain research topics and empirical sites of practice configured each of the assemblages in certain ways. For example, methods assemblage for exploring encounters with data representations primarily allows to study everyday and professional practices of laypeople. The methods assemblage for tracing dynamics of data infrastructures in my sample, is, in contrast, constructed from empirical research about marginalised populations, often outside of European borders and serves to understand implications of datafication processes on these populations. The methods assemblage for reconstructing datafied regimes led me back to the European and broadly Western perspectives on datafication processes enacted through politics and governance. Each methods assemblage is constructed around means through which datafication processes are enacted, such as data representations, data infrastructures, or datafied political and governance regimes.

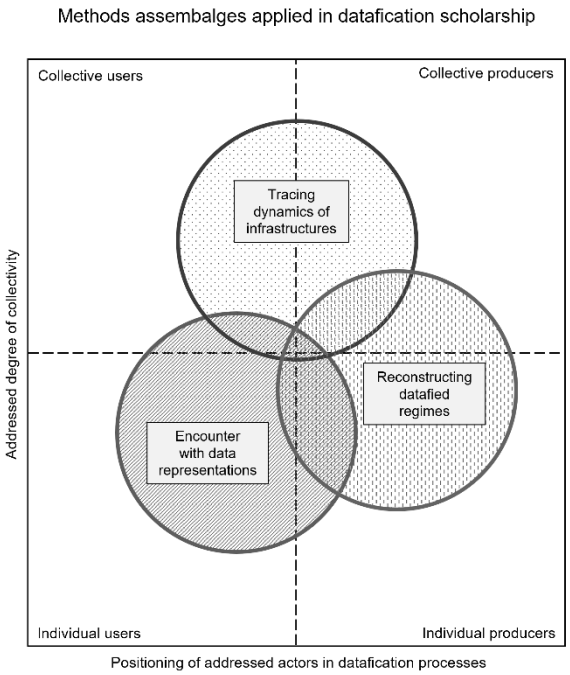

Figure 6-1 The continuum of methods assemblages for data inquiry

Figure 6-1 illustrates how these three methods assemblages can be placed within the heuristic developed in chapter 5. So, scholars studying encounters with data representations primarily focus on individual people who do not negotiate these representations themselves, but rather experience and enact datafication processes upon encountering these. These people—both individuals and collectives—translate encountered data representations according to their understandings, motivations, and affectivities in order to situate these representations in their lived realities. Datafication scholars, then, are interested in lived experiences and perceptions in which such a translation and enactment of datafication processes takes place. Research projects in which methods assemblages for exploring encounters with datafication processes are enacted, provide conceptualisations of datafication re-situated in the lived experiences of people and communities, sensitive to their values, needs, and goals. In my sample, several scholars enacting this methods assemblage conducted participatory studies that allowed them not only to create new encounters with data representations for their study participants to learn about them, but also to learn together and from the study participants, as I14 explained in one their quotes referenced in section 6.2. The methods assemblage for studying encounters with data representations, therefore, can be placed in the lower left square of the heuristic, as figure 6-1 illustrates. Although the methods assemblage is primarily placed in this square, indicating its focus on individual people affected by datafication processes, this methods assemblage also serves to build sensitivity to the various kinds of agency actors encountering data representations exert. The actors addressed with such a methods assemblage, therefore, can be differently positioned on the continuum between 'users' affected by datafication processes, and 'producers' holding power to negotiate data representations in the first place. As one of the interviewed experts, I19, argued, the distinction between 'users' and 'producers' is an analytical one, while in practice it is difficult to disentangle both in recursive enactment of datafication processes. The methods assemblage for exploring encounters with data representations, therefore, provides a foundation for moving in the mirky waters between 'usage' and 'production' of data, particularly when actors chose not to act upon their encounters with data representations or attune their everyday and professional activities—exert their agency in different, mundane ways than already better explored practices of resistance or activism.

The methods assemblage for tracing dynamics of data infrastructures tackles this challenge by addressing actors in positions to enable datafication processes who design, develop, operate, maintain, and overwatch data infrastructures as well as actors who utilise these infrastructures. Both

kinds of actors are understood as collectives, for examples states and public administrations or technology providers who develop and operate data infrastructures, or individual people as, for example, (non-)citizens—representative for a bigger population. These two kinds of actors are related to each other through data infrastructures and digital data moving across them. The movement of data is enabled by actors' work 'behind' the data infrastructures, through which certain actors are put in relation to others or excluded from relations. To understand these inclusions and exclusions, scholars applying methods assemblage for tracing dynamics of data infrastructures conduct long-term research at the relevant sites of practice and gather tacit knowledge about various political, social, economic, and cultural aspects of that site, also by building relationships of mutual trust with relevant stakeholders. By being able to draw on such tacit knowledge and by attending to the relations between different actors created through data moving across data infrastructures, datafication researchers, then, can understand the implications of datafication processes for various actors. This methods assemblage, therefore, can be placed in the middle of the upper half of the heuristic matrix I constructed. Such placement illustrates, first, how, tracing dynamics of data infrastructures, different kinds of actors can be addressed, both utilising these infrastructures, and developing, providing them. This placement also reflects predominant analytical attention to collective actors, even though some research techniques applied within this methods assemblage address individuals as representatives of a certain actors' group or a population. The examples discussed in section 6.3 range from individuals as citizens, individuals as professionals such as teachers, or individuals as members of other groups, such as e.g. learners. Putting forth the collectivity of the actors examined with the help of the methods assemblage for tracing dynamics of data infrastructures opens spaces for reflection about relations such datafication research creates: for example, relations of these individuals to the groups they are attributed to by datafication scholars. Before I continue discussing this further in the next chapter 7, I make concluding remark about the third methods assemblage.

The methods assemblage for reconstructing datafied regimes allows to explore practices of similar actors enabling datafication processes as the methods assemblage for tracing dynamics of data infrastructures. Reconstructing datafied regimes also means investigating work practices of public and political actors, or technology providers. The focus, however, is not on the work these actors do to relate themselves and other stakeholders to each other, but rather on their negotiations of data representations that are, then, put on the move across various socio-technical systems. The products of these negotiations are inscriptions made about datafication processes and digital data into the products of the actors' work, for example technology at large or software and algorithms more specifically, and policy documents when political and public actors are studied. Analytical attention of researchers applying this methods assemblage, therefore, is directed at the ways in which the results of these negotiations of data representations are ultimately inscribed in technologies or policies that enact particular datafied regimes. To reconstruct these regimes, datafication scholars require information about technology and policy design processes, involved stakeholders, their diverse interests, and internal conflicts emerging from such heterogeneity. Often, such information is not disclosed to the researchers due to the proprietary character of the products developed by technology providers, or due to politically sensitive issues of such an inquiry. Common for the methods assemblage for reconstructing datafied regimes, therefore, are alternative ways to investigate how data representations were negotiated, without direct access to the relevant stakeholders or documentation. Some of the scholars enacting this methods assemblage approach these datafied regimes through examining software interface and affordances, or by analysing algorithms. Others investigate policy documents in order to reconstruct the outcomes of the negotiations about data representations used in these policies. Interrogating datafied regimes in these ways allows to question, how and what meaning is attributed to certain data representations and the actors subject to these datafied regimes. Actors negotiating datafied regimes make attributions about meaning, while algorithms and policies they produce put these attributions to

work. These attributions, then, are further distributed, creating relations between various actors within and across socio-technical systems. When stakeholders subjected to these datafied regimes through policies or simply their use of certain algorithms encounter these representations, they translate these according to their own lived realities and re-enact them. In this way, all three methods assemblages are related to each other in an ongoing process of enacting and re-enacting datafication by various actors.

My analysis in this chapter shows that the three constructed methods assemblages are enacted by datafication researchers to seek different kinds of knowledges about datafication processes. Both methods assemblages and kinds of knowledges about datafication they help to acquire are distinctive, but in empirical research, they overlap depending on research questions and empirical sites of practice. Figure 6-1 shows how methods assemblages overlap according to the sample of analysed interviews. Some scholars conduct research that is predominantly matching with one of the three constructed methods assemblages: for example, I27's study on datafication and education can be primarily read as one enacting methods assemblage of reconstructing datafied regimes. In other empirical studies, elements of two or three methods assemblages can be found, particularly when interviewed experts discussed multiple research projects conducted iteratively in order to explore datafication processes from multiple perspectives, such as I4; I17, and I19 whose research is mentioned in all three sections of this chapter dedicated to individual methods assemblages. Exploring these empirical studies with the help of the heuristic I developed allows to reflect on different means through which datafication processes have been enacted, different kinds of actors involved in this enactment, and the relations between all these that ultimately produce a re-situated conceptualisation of datafication in empirical social research. As I will discuss in the next section, such a reflection should provide a springboard for further, different, and hopefully more generative knowledge production about datafication and its critique.

As my analysis indicates, to produce such re-situated conceptualisations of datafication processes, interviewed experts conduct empirical research that starts with "a hunch" (I24, pos. 6) as a quote in the beginning of this chapter states. Such a question-driven approach allows scholars to investigate datafication processes as situated in the lived realities rather than defined based on certain methodological (e.g. discourse analysis) or philosophical (e.g. structuralism) assumptions about what datafication should be. A research idea and, later, planning and design of an academic inquiry about datafication beginning with a 'hunch' means that researchers are interested in understanding some specific aspect in the multitude of aspects relevant to datafication processes. Then, besides the heuristic assembling methodological approaches in data studies I developed and illustrated so far, I show that different kinds of knowledges sought by researchers about datafication processes are central for constructing and identifying different methods assemblages. My analysis of the sampled literature and the interviews with the authors as well as some further datafication experts indicates that there are three main kinds of knowledges sought in empirical studies on datafication processes, ranging from 1) understanding of lived experiences and perceptions of datafication processes, to 2) their socio-political, historical, economic, and cultural implications, to 3) the understanding of negotiations of data representations in decision-making. My analysis also illustrates that these kinds of knowledges are intertwined with each other. They are not acquired by addressing certain kinds of research objects, for example only text or only visual material, or only experiences and practices. Rather, in research such as the projects discussed in my analysis, research objects, research techniques required to approach them, and other elements of a research process are assembled together in accordance with an empirical enactment of datafication processes. In the next chapter, I elaborate on these kinds of knowledges in more detail, relating them to the methods assemblages discussed here and returning to the arguments about academic knowledge production and critique made in the theoretical chapters of the thesis.

7 Data studies: what is at stake?

One of the interviewed experts shared following view on the current state of data studies:

> "I feel like, when I think about conferences and broader academic forums where I've been listening to a lot of presentations on data and having a lot of like, I guess critical data studies, I actually feel like we're kind of stuck. I feel like we've gotten to a point where we're able to critique, […], we are able to identify the stakes and why datafication can lead to massive societal problems. But it seems- those conversations seem to end up like an impasse between scholars pointing out all these issues and capitalism. Like companies doing this, owning the technology still data mining and data mining been a massive part of our economy now. And so, it just kind of ends up being like a shrug, like what are you going to do?" (I21, Pos. 23).

The 'impasse' formulated by the expert cannot be solved easily and my thesis does not attempt that: much more and probably different kinds of research as well as more substantial systemic change would be needed. The quote, however, makes clear that data studies need to continue being critical, but at the same time the critique they produce requires rethinking. Providing this statement in the beginning of this chapter, I rather aim to illustrate the stakes of (not) reflecting on data studies as a research field. This chapter of my thesis is concerned with the issue of critique in data studies: what is the role of critique and how reflecting on data studies with the help of the heuristic I developed can advance critique in new ways?

Common for data studies is an understanding of datafication as a recursive process in which relations between digital data and social realities are established and these social realities are constructed (e.g. Couldry & Hepp, 2017). These relations are socio-technical (e.g. Jarke & Breiter, 2019) as they are emerging based on technological transformations and affect involved stakeholders in different ways, as these technological transformations are, to a great extent, enabled by commercial companies profiting from the technological opportunities to gather and recombine big amounts of digital data at a high speed in order to re-use these data for commercial purposes (e.g. Gerlitz & Helmond, 2013; Gillespie, 2010). Some scholars highlight how datafication processes are historically embedded, for example as they follow colonial patterns (e.g. Lauriault & Taylor, 2019). Some of these conceptualisations underscore technological transformations underlying datafication processes, such as the ability to gather and process digital data fast, others, in turn, focus on how social relations are reshaped through datafication processes, for example as certain actors acquire more power than others. Yet other scholars offer critique to these currently dominant understandings of datafication (also see Bates, 2022 for extensive discussion of various lines of critique). For example, some notice that such understandings neglect everyday mundane situations and relations in which datafication processes are being re-enacted continuously by laypeople (e.g. Kennedy & Hill, 2018) or how these laypeople are conceived of as being ignorant, gullible, and lacking any agency in academic discourses (Livingstone, 2019, p. 171). Some researchers elaborate on the questions "[c]ui bono?" and "[w]ho owns the means of knowledge production?" formulated by Star (Star, 1995, p. 3 both). For example, Whittaker (2021) addresses the influence of big tech industry on the research and academic development of machine learning algorithms and academic discourses about opportunities and challenges these bring about. Such critique, however, builds only one part of critical discourses in data studies. More dominant critique is directed at 'identifying the stakes', the risks, and problems of datafication processes, as I21 points out during the expert interview quoted above.

These dominant kinds of critique shed light on individual cases and their implications, for example specific algorithms (drawing on a longer history of such critique, e.g. Introna & Nissenbaum, 2000) as well as their implications for certain groups of people such as People of Colour (e.g. Benjamin, 2019; Noble, 2018), specific companies and their products such as social media platforms (e.g. Bucher, 2018), or specific methodological techniques and the ways these are presented and imagined by different actors (e.g. Heuer et al., 2021). This is not to say this kind of critique is not important anymore: on the contrary, the abundance of discriminatory, eugenic, and harmful technologies that continue to hit the markets all over the world require further attention and interrogation. As the examples listed above illustrate, such critique is also attentive to the arguments of feminist technoscience scholars like Haraway (1988) and Suchman (2002) who argue that objectivity of knowledge production can only be achieved locally, by situating research in the context of its production. It means, being reflexive about the "agential cuts" (Barad, 2003, 2007) made in research processes and what these cuts include, exclude, and other. It also means being reflexive about the ontological politics of these in-/exclusions and otherings. This is exactly what this kind of critique does, pointing out to the inclusions, exclusions, otherings, (in)visibilities, or discriminations enacted in datafication processes. The stakes, risks, and problems identified there, however, not only persevere but also are often known to the actors enabling datafication processes such as technology providers and instrumentalised by them, for example by shaping "a positive narrative in response to growing regulatory and public pressure, alongside the industry's clear willingness to silence and punish critics" as Whittaker (2021, n.p.) discusses in regard to AI research.

The challenge this critique leaves the field of data studies with, therefore, is how to reflect on the academic knowledge production often intertwined with and relying on the very same techniques, infrastructures, and regimes. Beyond this reliance on the resources of actors criticised in data studies, academic critique can be increasingly read as holding to account actors discussed in these individual critical studies or at least raising public awareness about their, sometimes highly problematic, practices and products. Further, as I18 mentioned in their interview, critical, left-leaning academics often find 'what they are looking for' as they find the perils of datafication processes where the political left find the perils of capitalism. It is also by no means to say that such kinds of critique or taking political positions in datafication research in social sciences are unimportant. Rather, it raises the question of how to acknowledge and reflect these positions and political, social stances in academic knowledge production for data studies. In different domains and fields of social sciences, different answers have been developed. Some suggest new concepts that reflect moral values, for example of justice (see Dencik et al., 2019; Heeks & Renken, 2018; L. Taylor, 2017 on 'data justice'); others conduct participatory and activist research that engages with stakeholders at the empirical sites of practice (e.g. D'Ignazio & Klein, 2020; Kennedy et al., 2015), yet others propose alternative methodological approaches to social research altogether, for example by introducing a feminist notion and ethics of 'care' into discourses about datafication (e.g. Baker & Karasti, 2018; Fotopoulou, 2020; Lupton, 2020; Mattern, 2018; Zakharova & Jarke, 2022) and academic inquiry (e.g. Law, 2021; Law & Lin, 2020). My thesis is also largely built on the arguments developed by feminist technoscience scholars like Donna Haraway, Karen Barad, and Annemarie Mol, thus, I extend on this notion of care in my discussion of practices of critique and reflection in data studies.

As discussed in chapter 2, Latour (2004) proposed the notion of matters of concern for addressing critique as practice of situated knowledge production. These concerns reflect the society, social practices (p.245), they are "a multifarious inquiry […] to detect *how many participants* are gathered in a *thing* to make it exist and to maintain its existence" (Latour, 2004, p. 246 original emphasis). A matter-of-concern academic critique, for Latour, is about offering "the participants arenas in which to gather" (ibid.). While this seminal critique of academic critique is not aimed solely at debunking ideas but also at 'assembling' these, the concept of matters of concern has

become an object of critique itself. Especially feminist scholars advancing feminist concepts of care have noticed that a matter of concern lacks affective and ethical, normative connotations that are important in producing situated, local knowledge about technologies (e.g. Puig de la Bellacasa, 2011). So, Puig de la Bellacasa (2011) describes the differences between concern and care as a difference between thoughtfulness about a certain topic and a practice dedicated to change:

> "Understood as affective states, concern and care are thus related. Care, however, has stronger affective and ethical connotations. We can think on the difference between affirming: 'I am concerned' and 'I care'. The first denotes worry and thoughtfulness about an issue as well as the fact of belonging to those 'affected' by it; the second adds a strong sense of attachment and commitment to something. Moreover, the quality of care is more easily turned into a verb: to care. One can make oneself concerned, but 'to care' more strongly directs us to a notion of material doing. Understanding caring as something we do extends a vision of care as an ethically and politically charged practice, one that has been at the forefront of feminist concern with devalued labours." (p.89-90).

Such notion of care, therefore, speaks to the theoretical approaches I chose as a foundation for my thesis, as it helps acknowledge research and academic knowledge production as a practice directed at producing social change, rather than matter-of-factually recounting the stakes and challenges of datafication processes. The notion of care (also as discussed in chapter 3) reflects the challenges data studies currently face both in terms of e.g. research ethics, mentioned in chapter 6, and academic critique produced in data studies.

Following this, the question, then, is not only how to make one own's position as a researcher visible and understandable for other scholars, but also how to acknowledge and reflect upon these multiplicities and stances taken by academics to understand datafication in certain ways? In my answer to this question, I join I18 who argues for more reflection on the performativity of methods in data studies.

> "I do think that in this, like emerging field that sometimes calls itself data studies, there isn't much reflection on how methods shape findings. And I don't know if you are aware of the social life of methods literature, John Law and people like that. It's very interesting how that hasn't made its way into data studies" (I18, Pos. 11).

Rather than viewing methods performativity—'how methods shape findings'—as a challenge to empirical research processes, I take it as a foundation and a starting point for constructing the methods assemblages and the heuristic for analysing these. Methods assemblages gather different elements of academic research processes, situating these in the empirical site of practice and relating research to the empirical practices in which datafication processes are enacted. An example for an analysis taking a similar position towards methods performativity provides Thompson (2020), arguing for 'speaking with things' as a methodological pathway for

> "engag[ing] with more performative methods in order to 'speak with things' to examine both how data is 'part of the making and shaping of bodies' and how 'the body [is] a site of data politics' (Ruppert et al., 2017: 6)" (p. 3).

Thompson suggests three such methodological pathways:

- "attuning to and becoming with data" as a description of "the social and material relations between human bodies and data fragments lends itself to analysis that considers how particular data come to matter – or not – and to whom" (p.4);
- "making data physical" in order to allow the "juxtaposition of data fragments and annotations create an opportunity to engage reflexively and critically with how women might speak with and through particular representations of their data-bodies to wider publics" (p.5);
- "changing data narratives": "[t]he growing elisions and augmentations of data and human bodies invite researchers to consider data activism as research methodology. How might

> people speak with, through and as data-bodies to change narratives and propose alternative data imaginaries that shape and become part of digital heritage practices and artifacts?" (p.5).

These three methodological pathways proposed by Thompson (2020) foreground different kinds of relations in which datafication processes are enacted and how explicitly addressing these relations makes a turn from identifying high stakes of datafication to more generative/productive engagement with the role of data and datafication processes in lives of different actors (in Thompson's analysis—women). Taking methods performativity as a starting point for data studies, similarly, can allow scholars to engage productively with enactments of datafication processes and go beyond the questions of bias in research techniques, instruments, and samples. With the methods assemblages and the heuristic constructed throughout my thesis, I aim to provide researchers in data studies with an analytical tool for reflection and situating not only of datafication processes in the empirical sites of practice, but also for situating research and critique in the site of knowledge production.

Extending on this argument of generative, productive role of methods' performativity in empirical research, we can revisit the above formulated challenge which the current lines of critique in data studies leave this field with. I have shown in my analysis so far, albeit I eschew using the term '*critical* data studies' in my thesis, critique is one of the core modes of knowledge production in data studies. The challenge lies in reflecting on the relationship between academic knowledge production about datafication and the very same/similar empirical datafication processes. Taking this as an analytical starting point and drawing on the argument about productive role of methods' performativity, my answer to this challenge is in reflecting on methods assemblages and practices of holding these together. In chapter 5, I elaborated on how empirical datafication research produces re-situated conceptualisations of datafication; in chapter 6, I analysed which methods assemblages enact these re-situated conceptualisations. Reflecting on these re-situated conceptualisations of datafication with methods assemblages, therefore, can help tracing practices of academic knowledge production (and in case of data studies—predominantly critique) back to the multiple elements of these assemblages. These elements include, 1) the researchers and their positionings in the academic field(s) and at the relevant empirical sites of practice, 2) the researched (persons and things) in their historical development, 3) particular research processes, practices, and procedures guided by the relevant research politics, and 4) the empirical site of practice in which research is distributed. In the core of such reflection are the positionings of various elements of the methods assemblages and their relations to each other. As D'Ignazio and Klein (2020) notice in their discussion of a project Our Bodies Ourselves (https://www.odbproject.org), in sharing knowledge between scholars and communities affected by datafication processes, this kind of collective knowledge production and

> "co-liberation require[…] not only transparency of methods but also *reflexivity*: the ability to reflect on and take responsibility for one's own position within the multiple, intersecting dimensions […]" (p.64, original emphasis).

My goal here is to show how reflecting with the heuristic and methods assemblages I developed helps to reflect on different kinds of positionings of research objects and of research interests in their interrelation.

My analysis embraces the impossibility of a unified understanding of datafication or universal datafication processes. Rather, it allows acknowledging explicitly how the studied datafication processes are situated empirically. The methods assemblages and the heuristic for their analysis I constructed based on my synthesis of literature and interviews present my take on methodological reflection in data studies. The methods assemblages placed within the heuristic matrix visualise core categories alongside which empirical datafication scholars make analytical 'agential cuts' that determine distinctions between the research and the studied practice locally both within the empirical site of practice and within the site of knowledge production. Continuing the metaphor of 'cuts' proposed by Barad (e.g. 2003), in chapter 3 I also drew on Knorr-Cetina's (2002) notion of a 'dissecting room', in which—in her example microbiologists—were detaching their

objects of study from real bodies of laboratory mice and cutting them into pieces to make these objects of study analysable. After such an analysis, the dissected parts cannot be put back together into a Frankenstein mouse but find their place and become situated anew in the social reality of the laboratory and beyond. In chapter 3, I suggested that data studies operate in *somewhat* similar manner, conducting its own dissecting of datafication processes. In contrast to the working practices in a microbiology laboratory, the elements and aspects of datafication processes turned into objects of study do not cease to exist at the original empirical site of practice but become re-assembled as a part of the datafication scholar's methods assemblage. It is this analytical process of dissecting/detaching and re-situating that the methods assemblages and the heuristic for their analysis, developed in my thesis, reflect.

In line with the relational view on datafication processes shared in data studies, my findings show that datafication can be understood as a multiplicity of processes, at the same time enabled, enacted, and experienced through the relations between individuals, collectives, data and their representations, infrastructures, and algorithms, or policies and the translations these different actors undertake as they render themselves and each other into digital data and back. Furthermore, as my analysis suggests, datafication research is interdisciplinary and, according to several interviewed experts, also requires interdisciplinarity in order to investigate datafication processes from different perspectives (e.g. I6, I19, I29, I25, I29). These different perspectives on datafication processes co-exist with one another. My findings presented in chapters 5 and 6 indicate that the 'cuts' datafication scholars draw within these relations and translations are situated in each studied empirical site of practice, rather than driven by conceptual definitions of datafication. Thus, empirical research on datafication produces new, re-situated conceptualisations of datafication processes under study. Such a perspective on datafication processes acknowledges their 'ontological multiplicity' (Law, 2004; Mol, 2002), but also highlights the multiplicity of theoretical and disciplinary concepts required to explore datafication processes. Both for Mol (2002) and Law (2004), empirical research acknowledging ontological multiplicity should focus on the relations in which each of the multiple realities is situated.

As I have argued throughout the thesis, data studies face various methodological challenges, some already widely accepted and discussed in social sciences, for example within the 'double social lives of methods' debate (Law et al., 2011; Law & Ruppert, 2013; Ruppert et al., 2013). These topics primarily concern the implications of methods' performativity, objectivity of academic research, aims of academic critique, and methodological reflection. Other challenges emerge alongside technological transformations that allow researchers to rely upon information systems and infrastructures built by actors outside of academia for non-academic purposes, such as website APIs, social media data, and instruments for their analysis. These challenges are specific to the object of study of datafication research and often concern computational, digital techniques of data collection and analysis. Datafication processes are complex, as they are continuously enacted and re-enacted by different stakeholders. Datafication processes are relational and recursive, as they transform social realities rather than solely represent them. Studying datafication processes empirically, then, means dealing with these complexities, multiplicities, and recursivities. Grounding my analysis in the theoretical approaches to methods performativity and academic discourses of the double social lives of methods, I argue here that academic knowledge production about datafication should not be seen separate from the methodological approaches, but rather in concert with these. This is why, methodological reflection on how datafication processes are studied empirically is an important part of academic knowledge production.

The methods assemblages I constructed allow such a reflection as they are organised around the answers datafication scholars give to the question of what we are talking about when talking about datafication and how can we study datafication processes. The methods assemblages give answers to this question, highlighting either 1) encounters with data representations, 2) dynamics of data infrastructures, or 3) datafied regimes as means through which datafication

processes are continuously enacted and re-enacted. This listing, while not exhaustive, suggests reflecting on what 'datafication' stands for in a situated way in each empirical research project, what means help enact datafication processes (particular kinds of technologies, policies, or something else), what is addressed as data in this particular research and how different kinds of these 'data' encountered by researchers relate to each other locally. These questions arise from taking methods' performativity as an analytical staring point. Bridging the concept of methods' performativity with the feminist concept and ethics of care, as I propose here, reframes reflection on performativity of methods in data studies to a specific practice of more care-ful research and an ethico-political obligation (see Puig de la Bellacasa, 2017 for defining care as such an obligation). My goal here is to address this more practically, by developing a heuristic that provides categories for reflection. These categories are sensitive to the multiple relations and positionings enacted in empirical datafication research.

With the methods assemblages and the heuristic matrix for their analysis constructed in my thesis, I illustrate different kinds of sensitivities required to empirically explore and conceptualise datafication processes in situated ways. By introducing the feminist concept of care, I react to the arguments made by datafication scholars who participated in my interview study. I showed that multiple interviewed researchers perceive the growing field of data studies as being 'stuck' (I21) in matter-of-factual critique of high stakes and risks of datafication processes, while more generative approaches to critique are yet to build. The heuristic I constructed illustrates how data studies themselves already work with some concepts and instruments that could lead to building such critically generative approaches. To these, I add established sets of sensitivities developed by feminist scholars in negotiations of 'care' as an analytical concept to extend on the rich and long-standing body of work that engages with invisibilities, exclusions, and otherness—processes central to datafication, but not new to social research. I believe, the notion of care and feminist, care-ful, care-related methodological vocabulary fits well with the tasks of future, critical research on datafied societies—maintaining relations between data and societies, establishing and continuing generative, productive practices of digital participation and heterogeneity of data representing various communities, repairing the damage invasive, extractivist datafication processes have already caused to individuals, communities, and the society at large.

Overall, my analysis calls for developing new vocabularies that are not connoted with long-standing binaries between society, technology, culture, and nature, research and empirical practice, methods and knowledge. The three methods assemblages I constructed are, instead, allocated along the continuums reflecting the degree of collectivity of actors addressed in empirical datafication research and the positionings of these actors in the processes of datafication between data 'use' and data 'production' poles. Each of the methods assemblages, then, allows scholars to seek particular kinds of knowledges about these datafication processes. By explicitly addressing different kinds of knowledges and the categories such as extent of collectivity of addressed actors alongside which analytical 'cuts' are made, the heuristic constructed here also allows reflecting on how knowledge about datafication processes is produced and 'who owns the means of production' (Star, 1995).

Figure 7-1 draws on the findings presented in chapters 5 and 6 and illustrates what kinds of knowledges are sought in empirical datafication research according to my sample and how these knowledges match with the constructed methods assemblages. So, scholars exploring encounters with data representations are interested in understanding lived experiences and perceptions of people—mostly individual laypersons—of data representations in their everyday and professional lives. Scholars tracing dynamics of data infrastructures are interested in understanding socio-political, economic, and cultural implications of datafication processes enacted through the movement of data across these infrastructures. Finally, scholars reconstructing datafied regimes aim to understand how data representations are negotiated by the actors in positions to enable datafication processes. Even though I identify these three kinds of knowledges based on the three methods assemblages presented in this chapter, they do not exclusively match with each other and,

as my analysis illustrates, many datafication scholars seek answers to multiple questions in their empirical research projects. Furthermore, the way methods assemblages are constructed around certain means of enacting datafication processes allows much flexibility in defining what a scholar aims to understand about datafication. For example, an encounter with data representations can be addressed analytically as an affect, a perception, a practice, and more. Similarly, by attending to data movement across data infrastructures, not only who and what moves data can be addressed, but also temporalities and topologies of the movement required to understand what implications this movement bears for involved stakeholders (e.g. I5, I24). As my synthesis illustrates, datafied regimes can be negotiated not only by political actors, but also by technology providers and multiple intermediaries between the decision-makers negotiating data representations and the actors subject to these regimes.

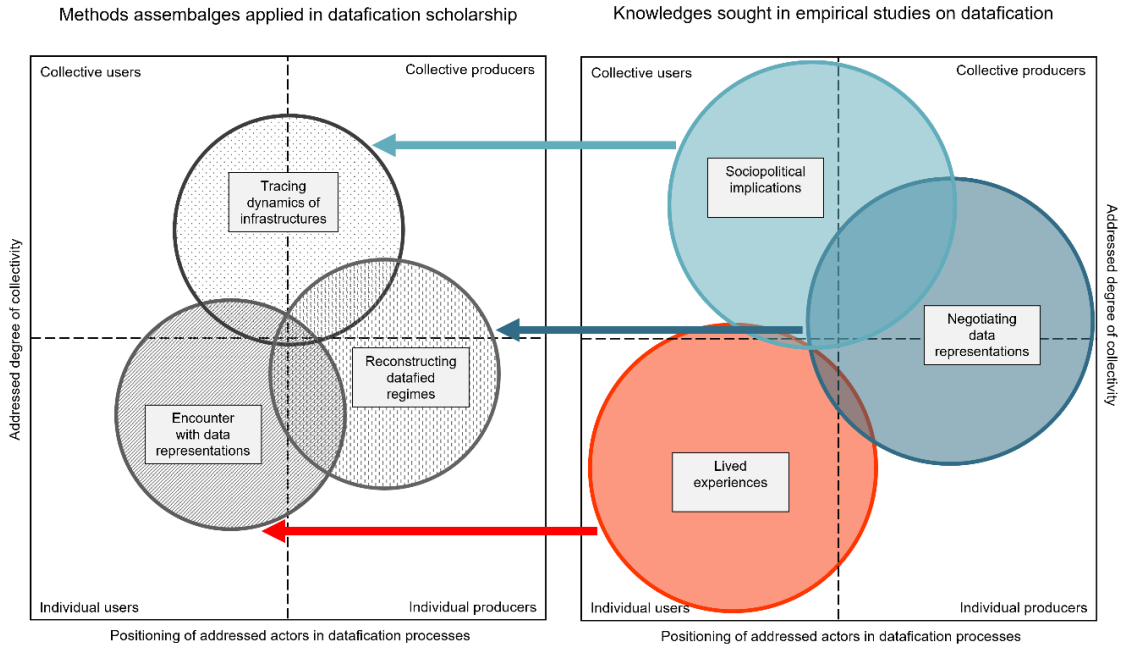

Figure 7-1 Methods assemblages and the knowledges sought with the help of these

Figure 7-1, however, is in itself a visual representation, a research artifact, the role of which in qualitative research deserves its own analysis, albeit it is out of scope of my research presented here. To briefly summarise, visualisations like the one discussed here have been discussed in relation to theory and theorising (Weick, 1995), to mapping, analysing, conceptualising, and communicating empirical research (e.g. Langley & Ravasi, 2019). Rather than being free from various assumptions, visualisations in qualitative research illustrate the analytical 'cuts', made by their authors. Moreover, the arrows and boxes used in visualisations follow their own conceptual and aesthetic conventions and are themselves performative (Langley & Ravasi, 2019). Figure 7-1, thus, illustrates the analytical 'cuts' I made during my conceptual and methodological inquiry into datafication research and provides a modelled, simplified picture of the methods assemblages I identified in the analysed sample of academic publications and interviews with their authors. For example, the choice of circles as a form for representing the methods assemblages, marks one of such analytical 'cuts'. Similarly, the equal scale of the circles labels another 'cut', as for my argument here a quantification of cases in which each of the three methods assemblage were enacted is less relevant than (and hardly possible due to) the overlapping between the methods assemblages in many of the analysed empirical studies on datafication processes.

Presenting this heuristic with the help of two continuums has advantages. For example, as mentioned in the concluding section of the previous chapter, it allows to position actors addressed empirically not as either data 'users' or data 'producers', but as being affected by datafication processes and having agency to re-enact them, even in mundane settings, at the same time. As

discussed in chapter 6, methods assemblages overlap and complement each other, while there might be other methods assemblages, not covered by my sample. The methods assemblages, although being distinctive in regard to their central elements such as means through which datafication processes are being enacted, extent of collectivity of addressed actors, and their positioning in the datafication processes, relate to each other rather than they are self-excluding. Each of the three identified methods assemblages can be understood as an empirical investigation at a certain point in the enactment of datafication processes and all three can be directed at studying a similar empirical site of practice. The aim of the remaining section of this chapter is to 'apply' the heuristic for analysing methods assemblages developed here and illustrate how using this heuristic as a reflexive tool opens pathways for different kinds of critique. I formulate these pathways as questions, although my goal here is not to answer them, as they would require empirically situated answers. Rather, I outline some of these pathways in the following, as I discuss various kinds of relations that the developed heuristic and methods assemblages shed light upon.

First of these relations are depicted in my heuristic as two continuums between various kinds of actors—individual and collective ones—and their positionings as 'users' or 'producers' in enacting datafication processes. Questions about relations between collectives and individuals they comprise of are methodologically relevant, as they force reflection on the applied research techniques. Not only do different research techniques represent collectivity differently (compare, for example, a network graph of log data to an interview with one or several representatives of a collective). My analysis presented in chapter 6 indicates a disproportion between studies directed at individual actors (for example laypersons encountering data representations while reading a newspaper) and studies directed at an exploration of datafied 'logics' and regimes negotiated by such collective actors as technology providers or governments and public administrations. According to my analysis, both the former and the latter studies can resort to conducting interviews with individual people. While this is an established and by no means unproductive form of exploring collective actors, readers of such research require understanding about how these interviewed individuals are situated within the collective. Individuals within such collective actors, while bringing their own perspectives, motivations, goals, and values, can be bound to lesser or greater extent by shared collective goals, e.g. depending on the level of institutionalisation of such a collective actors (e.g. local community VS international corporation). Moreover, reflecting on collectivity of different actors involved in enacting methods assemblages of empirical datafication research also can mean reflecting on the positionings of researchers in various collectives and on *research as collective practice*. Activist, action, and participatory approaches to academic research provide some examples of how multiple positions, interests, and politics are fused in and enact empirical studies.

These questions, while still relatively novel to data studies[23] have been extensively discussed, for example, within participatory research approaches (e.g. Bennett & Rosner, 2019; Costanza-Chock, 2020; Vines et al., 2013). This body of literature elaborates on research ethics in participatory research in a way that can be helpful for data studies, too: for instance, in reflecting on and planning empirical research designs, especially if the aim of such designs is 'empowerment' of study participants. For example, Robertson and Wagner (2013) propose to question whether study participants "actually have decision power" (p. 82) and what kinds of it; whether study designs "encourage participants' abilities to learn" (ibid.) or "guide designers and researchers to analyse and develop their interests and attitude towards participants" (ibid.). Further the authors urge to reflect on how to

> "deal with a justified loss or change of design [and research] focus, for example when participants identify problems that require non-information technology solutions while the

[23] Even though ethical questions raised here have been discussed in (feminist) data studies (see Cifor et al., 2019; Gardner & Kember, 2021; Kennedy et al., 2015; Posner & Klein, 2017).

> process was initiated to design [or study] information technology?" (Robertson & Wagner, 2013, p. 82).

These questions for reflection are not only relevant for data studies in exploring, for example, how individual actors from marginalised groups encounter data representations or are subjected to certain datafied regimes. Expanding on participatory research also allows to incorporate in data studies positions articulated by several interviewed datafication experts (as mentioned in chapter 6). These experts (e.g. I30) argue that a more differentiated analysis rather than utterly critical one is required to understand the roles of more powerful stakeholders—technology providers or political actors—in enacting datafication processes. The questions mentioned in this paragraph primarily address the relations between researchers and researched actors as active participants and collaborators in the collective practice of academic knowledge production. An overview of literature about datafication in chapter 2 of my thesis also illustrates a manifold of critical contributions reflecting on the relations between researchers and their digital, computational research instruments (e.g. Rieder & Röhle, 2017; van Es et al., 2021). These issues complicate academic knowledge production, but, if reflected upon, may provide a springboard for developing more generative and careful critique in data studies. A pathway for such critique emerging from reflecting on various relations in which data scholars are entangled is directed at data studies and researchers doing this work themselves.

Another kind of relations that requires attention in data studies is the relation between what can be described on the continuum between 'data use' and 'data production', also referring to the possibilities of stakeholders in acting upon data and datafication processes. As several scholars have pointed our recently, neglecting these relations leads to producing images of users as ignorant and lacking agency in the datafied worlds (e.g. Livingstone, 2019; Pruulmann-Vengerfeldt & Hörste, 2020). This regards not only individual 'users' as referenced critiques argue, but also different kinds of collective actors also engaging in enacting and re-enacting datafication processes, such as data activists, social movements, and local communities. For example, drawing on similar arguments recently made by Beaulieu and Leonelli (2021) and Braun and Kropp (2021), in relation to datafication processes (and more specifically to the processes of, often commercial, data collection) these users can be rather seen as data producers, while other actors engaged in processing and re-sale of these are these data's users. Beaulieu and Leonelli (2021) also identify data scientists and data managers, who, together with the data users and producers, in the outlined definition, constitute what the authors refer to as data workers (p. 11). Reflecting on the positioning of various actors on a continuum between data use and production opens a pathway for further discussions about who owns means of data production. Directing this reflection on the practices of data studies themselves also relates the issue of data use/production to the ethical questions of who owns the means of *knowledge production* and what role different actors play in that (take, for example, ethical consideration in research that relies on gig workers such as discussed by e.g. [34]). Further, such collective actors as, for example, technology providers, are broadly understood as being in position to enable datafication processes and with the change of their practices, new kinds of resistance or more production datafication processes could be enacted. My literature analysis presented in chapter 5, however, indicates that a number of datafication scholars have contested such view, noticing that both activists aiming at enacting datafication processes differently and commercial companies do not always have much agency to resist and, ultimately, continue the practices of data and value extraction they initially set out to oppose [5, 20, 49]. It is particularly important in light of dominant concepts about digital data and datafication processes, that conceive actionability upon data as one of the central ways to exercise agency in a datafied society. For example Couldry and Hepp (2017) refer to Amoore's (2011, p. 29) argument about data being rendered actionable in order to be put to use. While for some actors this might be true, such actionability of data might place additional pressure on other actors participating in datafication research. As an example reported by I12 illustrates, some of their study participants were not interested (at least immediately

during the research) in reflecting or changing their everyday and professional practices based on the visualisations of their app and media use. Rather, these study participants were interested in learning about themselves through the data, e.g. to support their current practices. It puts studies that aim to give back data or enable and empower study participants in their practices of enacting datafication processes in a more complicated position. Particularly regarding the last example, this pathway for critique in data studies is also a kind of a self-critique, aiming at reassessing how the research goals and knowledges sought by data scholars are one of the backbones for enacting the methods assemblages and how these assemblages are gathered with respect to these interests and goals.

In this chapter, I discussed the findings of my research in relation to the concept of methods' performativity as a central theoretical foundation of my study. Taking methods' performativity as a point of departure for the further analysis and bridging it with the feminist concept and ethics of care, I discussed the role of critique in data studies and sketched pathways for (self-)critique sensitive to this methodological performativity. The methods assemblages and the heuristic for their analysis developed in my thesis serve as a reflection tool for creating these pathways. Applying this heuristic, various questions for further research and reflection within the emerging field of data studies arise: how are affiliation and membership constructed and negotiated both through datafication processes and through academic knowledge production about datafication? How are these affiliations different to the attributions inscribed in technologies, data representations, and datafied regimes? (An example of empirical research that poses some of these questions can be found in my sample in an article by Taylor and Richter [47] reporting analysis of water supply systems in relation to the definition of citizenship in India.) Other question might touch upon the definitions and self-attribution of actors to their positionings on the continuum between data use and production; the issue of positioning non-human actors (such as technology) on this continuum also follows from the gaps left in the heuristic I developed. These and a myriad of possible other questions can be posed to reflect and situate datafication practices further in order to produce re-situated, empirically informed conceptualisations of datafication.

In regard to the feminist argument of care that I put forward here, the pathways of (self-)critique discussed in this chapter primarily reflect care as an ethical obligation, raising multiple questions of research ethics. This fits well with the overall goal of my thesis to conduct a methodological and conceptual inquiry into the emerging field of data studies and to understand better how methods assemblages help data scholars in producing re-situated conceptualisations of datafication. In this perspective on the field of data studies, care-fulness can be found in practices of reflection, which I advance with the methods assemblages and the heuristic for their analysis developed here. At the same time, as I have shown throughout my thesis, care can also be understood as a local practice of maintenance, continuation, repair, and tinkering. While this chapter was dedicated to discussing my findings—on the *field* of data studies—such practical understanding of care could not be implemented: after all, the reflection tool developed here does not serve as a practical guide but rather as an instrument for advancing sensitivities towards the complexity of datafication as an empirical process and an object of study. In the following chapter concluding my thesis, I take a peek into how this understanding of care as a practice in data studies can be realised and argue that besides reflection and advancing sensitivities for critical analysis, an overall care-ful methodological approach to data studies can pave a way forward for more generative engagements with datafication. As Law (2021) notices, such care-ful research exists already. In the concluding chapter of my thesis, I elaborate how adopting care-fulness in data studies opens different, more generative, albeit critical future directions for the development of the field.

8 A plea for care-ful data studies

In the introduction to my thesis, I quoted a feminist technoscience scholar Donna Haraway (2016) on the role of reflection in academic work and knowledge production, that she articulates so beautifully: "It matters what matters we use to think other matters with; it matters what stories we tell to tell other stories with" (p.12). My thesis is an attempt to understand and reflect upon what stories data studies tell to tell stories about empirical datafication processes. I explored the growing research field of data studies mapping out literature, visualising central research domains, topics, and methods, and analysing interviews with scholars who have conducted empirical studies on datafication in the past years. According to my synthesis, conceptual definitions of datafication applied by the authors of sampled literatures are reworked and re-situated according to specific research findings. These re-situated conceptualisations shed light on the kinds of means through which datafication processes are enacted at empirical sites of practice studied by datafication researchers. Around the means through which datafication processes are enacted empirically, I constructed methods assemblages of datafication research which encompass: 1) exploring encounters with data representations, 2) tracing dynamics of data infrastructures and data movements, and 3) reconstructing datafied regimes. The methods assemblages reflect not only what we are talking about when talking about datafication, but also different kinds of research interests assembled in research practices. So, each of the assemblages responds to one or several of the three kinds of knowledges sought by datafication researchers with their empirical investigations. These knowledges encompass an analytical interest in 1) lived experiences and perceptions of datafication processes by various actors, 2) social, political, historical, economic, and cultural implications of datafication processes on different stakeholders, and 3) negotiations and related tensions in defining data representations. Together, methods assemblages can be positioned within the heuristic for their description and analysis; they overlap and complement each other rather than they are self-excluding. This heuristic can be visualised as a matrix consisting of two continuums: 1) continuum describing the degree of collectivity of actors addressed by empirical datafication research and 2) continuum of positionings of these actors in the datafication processes between the poles of data 'use' and 'production'. I argue that together with the methods assemblages, this heuristic can be seen as a reflection tool for empirical data studies. This tool provides categories for reflecting critically on how research re-situates conceptualisations of datafication empirically and why, what role different elements of the methods assemblages have in it, and how are these elements involved in co-production of academic knowledge about datafication.

Even though I argued for using the term 'data studies' instead of 'critical data studies' in the beginning of my thesis, the notion and practice of critique remains central for the research endeavour of data studies scholars. Drawing on the discussion of critique in the previous chapter, I argue here that data studies require more care-ful, generative critique. So how does the concept of care in data studies relate to the methods assemblages and the heuristic matrix for their analysis developed in my thesis? I believe, both allow asking similar kinds of reflexive questions about the nature and means of academic knowledge production, about the relations between various actors at the sites of practice of knowledge production and empirical datafication research. The heuristic I constructed illustrates that data studies themselves already work with some concepts and instruments that could lead to building such critically generative approaches. The need for a care-ful critical approach in data studies becomes particularly visible in times of crisis such as the Covid-19

pandemic, when data analyses and (data-driven) decision-making is happening fast and the consequences of both can be substantial for vast parts of the societies (for examples of data studies with focus on the pandemic see e.g. Brunsdon, 2020; Micheli et al., 2022).

In the remainder of this concluding chapter, I propose my view on care-ful, generative critique in data studies drawing connections between the results of my analysis and most recent conceptual and empirical research on digital data and datafication, also sketching pathways for future explorations. First, following arguments put forwards by feminist technoscience scholars such as Haraway and Suchman, I understand as *care-ful and generative research that explicitly acknowledges situated positionings of researchers, their objects of inquiry, and other elements of methods assemblages in multiple social realities*. The sensitivities explicated through the notion of care can address different kinds of relations enacted in datafication research, such as relations between researchers and researched actors, researchers and their research instruments, between different kinds of actors involved in, encountering, working with, and negotiating data and their representations. The sensitivity of the notion of care to the affective, painful, troubling, and othered aspects of datafication processes has been discussed increasingly in social sciences. For example, Lupton (2020) argues that 'thinking with care'

> "generate[s] awareness of and attentiveness to the affective as well as social, cultural, and political dimensions of these assemblages. This approach can address questions that go beyond a preoccupation with institutional agency and the repressive nature of datafication and dataveillance, avoiding a techno-deterministic, top-down perspective that sets people and data/technologies as separate from and in opposition to each other." (p.3169).

Some scholars argue that such awareness can be best generated by 'slowing down' research practices (e.g. Mountz et al., 2015; Ulmer, 2017). While the debates about work practices in academia are not in the focus of my thesis and only emerge on margins of my discussion of the methods assemblages (albeit being their part), I believe that such slowing down also allows a more care-ful engagement with empirical datafication processes. For example, multiple interviewed scholars reported about challenges they faced in access to objects of inquiry, for instance in their studies of data-driven policy or various kinds of public and commercial technology providers. Some of the interviewees pointed out the need of waiting for developments on the site of empirical practice (e.g. pauses in research projects related to changes in political agendas). Reflecting on such challenges in terms of care helps viewing these as slowing down of research practices, viewing research processes as filled with interruptions and breakdowns. I believe, a critical generative approach could help to embrace these interruptions as springboards for further research: embracing these interruptions could mean engaging with the power relations, agency, literacy, materiality of datafication processes, their cost and financial, economic resources they require that become visible when research practices halt. So, Ulmer (2017), in a call for slow ontologies, suggests that such slowing down allows following "more sustainable rhythms of inquiry" (p. 206). While in Ulmer's contribution these sustainable rhythms refer primarily to natural cycles, I draw on this idea in relation to the temporal developments of empirical datafication processes. I believe, this is a useful notion for research that follows datafication processes in their rhythms and scale and is ready to accept the related 'interruptions' and what generally can be considered 'inefficient' or 'ineffective' or 'not working' in datafication processes alongside with the 'effective'/'efficient' elements and points in time in the development of these processes (see Bridges, 2021 as an example). Further, this care-ful notion of slowing down empirical research on datafication also bridges data studies with literature concerned with issues of temporality (e.g. Baygi et al., 2021; Coleman, 2018; Coletta & Kitchin, 2017; Decuypere & Broeck, 2020; Leavitt, 2019).

Second, *generative and care-ful critique in data studies is sensitive towards the manifold of relations (including affective ones) various human and non-human actors build in enacting datafication processes as well as in enacting methods assemblages for studying these*. I showed in my analysis how some datafication scholars already pursue some aspects of the care-ful research agenda, for example attending to the affectivities in relations with digital data. Another line of recently published literature focuses on so-called 'data professionals'—people whose work and labour are organised around generating,

cleaning, and maintaining digital data—across various research domains within and beyond social sciences such as medicine (e.g. Nordfalk, 2022; Pinel et al., 2020; Pinel & Svendsen, 2021), education (e.g. Lu et al., 2021; Whitman, 2020), in organisation research (e.g. Hockenhull & Cohn, 2021; Willems & Hafermalz, 2021), and further application domains of data science (e.g. Dencik & Stevens, 2021; Kaun et al., 2020; Kaun & Stiernstedt, 2020; Perrotta et al., 2022; Posada, 2022). Besides the focus on work relations, these studies also shed light on institutional and organisational perspectives on (doing) data. Other scholars, often also building on feminist literature, specifically explore how relations between human actors and digital data come to be and what these relations mean in the social realities of the affected human actors (e.g. Klumbytė, 2022; Lomborg, Langstrup, et al., 2020; Pinel et al., 2020). Another example of a care-ful engagement with the relations between actors enacting datafication processes provide Beetham and colleagues (2022) in their recent publication. The authors discuss a researchers' community under a hashtag #FemEdTech as a form of activism and collaboration/networking for articulating different voices. In that sense, such care-ful view on academic research also can be understood as an agenda for opposing some aspects of academic work. As I illustrated in the previous chapter, reflecting on empirical datafication scholarship with the heuristic and the methods assemblages I developed brings forth various kinds of relations enacted in these empirical methods assemblages. These relations are, then, incorporated in the re-situated conceptualisations of datafication processes.

Third, *care-ful and generative critique in data studies is not only critical, but also positive beyond the discussion of 'best practices' in the sense that it also shows positive, advantageous aspects of datafication prosses* (e.g. I3, I25). The many examples of datafication research in every chapter of my thesis showcase the pervasiveness of digital data in our social lives. With that, digital data and related datafication processes, for many actors in various situations in their personal and professional lives, are inevitable, and often essential. A care-ful academic critique of these datafication processes, thus, also acknowledges datafication processes as a starting point rather than an outcome of analysis, directing attention to the ways in which datafication processes can and cannot be lived through and coped with by the affected actors. Through the notion of care, both virtues and precarities of living with datafication can be addressed, drawing on how Hobart and Kneese (2020) discuss "radical care as a set of vital but underappreciated strategies for enduring precarious worlds" (p.2) that can be mobilised by different actors. So, Zegura and colleagues (2018) bring notion of care in relation to data science, discussing how care as "a process for making decisions and taking action that recognizes that facts and choices are value-laden and strives to keep those values present" (p. 8) is helpful in research collaborations with communities. As some of the interviewed experts underscore, it is however important to recognise that sometimes, reported best practices are only exceptional workarounds that only provide an illusion of a possibility of systemic changes (I3, I25).

In relation to these possible systemic changes, the fourth aspect of care-ful critique is central. So, *generative, care-ful critique takes into consideration normative and ethical aspects of data studies—not only addressing research ethics, but also the normative power of a research field in identifying its object and sets of methodologies for studying it*. On a more practical level, this ethico-political and normative aspects of care-ful critique bring forward socially desirable aspects related to datafication and the question of who decides how on what are 'desirable' datafied futures. For example, Fotopoulou (2020) elaborates on the notion of care in relation to data literacies and how this notion allows to develop normative principles for advancing data literacies and skills. Further, topics only briefly discussed in my thesis such as AI ethics and ethical design of ML-based systems build another line of research that is encompassed by this fourth view on care-ful critique in data studies. This research is currently widely debated in the domains of computer science, data science, and human-computer interaction (e.g. Birhane, 2021; Donia & Shaw, 2021; Gray & Witt, 2021). Others, such as Diaz-Bone at al. (2020) argue for emphasising normative aspects of datafication processes, for example how negotiations about data representations can be understood as justifications or benchmarking. This kind of critique is also about creating care-ful relations between datafication scholars and practitioners involved in enacting datafication processes as part of their professional practice. For

example, it concerns methodological literacies of practitioners who do work like data studies, for example in public institutions, NGOs, educational organisations, and in media production that is subsequently distributed among public decision-makers, actors in civic society, and broader publics. Such public reports and media coverage have a sustainable influence on civic and public understanding of datafication processes. As multiple scholars lamented after the latest series of whistle-blowing reports (see Sharp et al., 2021 for an example), these reports often 'uncover' what is already widely known and accepted in the scholarly communities. Similarly, besides written, textual reports of findings about datafication processes, recently an increasing body of work on the role of data visualisations has been published. These studies attend to visualisations both as a methodological, research-practical issue, and in relation to the politics of visual representations (e.g. Bowe et al., 2020; Fileborn & Trott, 2021; Zhao & Ye, 2022). If these reports and visualisations do not show enough methodological sensitivity to the many aspects of datafication processes discussed throughout my thesis, even though 'meant well' they could provide ever more findings inattentive to some exclusions/otherings and primarily attentive to others[24]. Collins and Pinch (2012 [1998]) argue, understanding science is important for "citizens who want to take part in the democratic processes of a technological society" (p. xv). Even more so is it important for people living in a datafied society, where data are pervasive, to understand how research on these data works and how knowledge about datafication processes is constructed. My thesis does not provide a full map of data studies and primarily develops a reflection tool for empirical scholars. By showcasing multiple intertwined elements of datafication scholars' methods assemblages, however, my thesis underscores the importance of reflecting on our research field methodologically, both for researchers and for those affected by this research. A more care-ful engagement with the matters of datafication research—also care-ful towards its readers within and outside of academia, e.g. by explicating authors' positionings in the field and research interests—might be one of the steps forward.

The sketched pathways together with the current scholarship already incorporating some of the enlisted research perspectives and practices constitute my plea for more care-ful research in data studies. As a reflection tool, the methods assemblages and the heuristic for their analysis developed here provide categories sensitive and sensitising to the variety of relations enacted in empirical data studies. These categories encompass means by which datafication processes are enacted, extent of collectivity and positioning of the affected actors within empirical datafication processes and, finally, kinds of knowledges sought by researchers who tie together each particular methods assemblage. These categories refer to the core elements of the methods assemblages that I identified in chapter 3: the researchers and their positionings in the field and the empirical site of practice; the researched; research processes and procedures; the empirical site of practice in which datafication processes are enacted. These categories, then, can be seen as an additional, alternative vocabulary for reflecting about research practices and their performativity towards academic knowledge production. As I discussed in multiple chapters of my thesis, individual methods as techniques and procedures of research data collection and analysis remain one of the building blocks of the methods assemblages but are moved from a central position to the periphery. This allows turning away from methodological discussions about methods (qualitative, quantitative, digital, computational, etc.) to the questions about what 'datafication' stands for empirically, what human and non-human actors are involved in enacting these empirical datafication processes, and how these heterogenous actors can be assembled in an empirical study and situated in their relations to each other.

[24] For an example of such dominant attention in media coverage to a certain group of people taking harm from their social media use see following letter to the Editor in the Wall Street Journal (26.09.2021) by Ysabel Gerrard, reflecting on media coverage of Instagram algorithms' harms to 'teenage girls': https://www.wsj.com/articles/facebook-instagram-mental-health-teen-girls-zuckerberg-11632426638

With that, I also join literature calling for use of multiple methods for studying datafication processes (see also Nikunen, 2021). In regard to the methods and techniques applied, my findings align with results of the literature synthesis presented by Flensburg and Lomborg (2021): in my sample, most contributions reported about qualitative research projects, while techniques such as interviewing and other ethnographically inspired methods such as observations were applied often. Multiple contributions reported findings of projects applying a mix of different qualitative techniques of data collection. Only a small number of sampled articles applied different kinds of digital methods such as a walkthrough or digital ethnography. Similarly, only a few authors applied computational techniques of data collection and analysis. While these findings align with other, systematic reviews of literature about datafication (see e.g. Flensburg & Lomborg, 2021; Özkula et al., 2022), they might also reflect the limited sample used in my analysis. For example, including empirical research from other research disciplines than social sciences could put these findings in question. So, scholars working in various research domains such as health, data studies, human-computer interaction develop new research methods and apply different kinds of techniques for their analysis, often combining computational procedures with qualitative and creative one in numerous new ways working on health, geography, doing cultural studies, applying computational methodologies for their research (e.g. Barry Born, 2021; Bleeker et al., 2020; Lupton & Watson, 2020; Tkacz et al., 2021; van Koningsbruggen et al., 2022). Yet others reflect on currently applied methodologies and techniques (e.g. Birhane & Guest, 2021; Leonelli et al., 2021; Nikunen, 2021). Following the construction of my literature sample, these research techniques find not much reflection in the methods assemblages. Finally, my findings demonstrate that in the analysed literature sample, rather little space was dedicated to the technological and hardware aspects of datafication processes. If they would, a further kind of a methods assemblage would be added to my findings, one that is probably organised around predominantly technological, hardware means of enacting datafication processes (such as physical objects and sensory devices), while these non-human actors would require positioning within the respective organisational/institutional actors designing, developing, and maintaining these. The question of positioning such non-human, technological actors on the continuum between data production and use could be interesting to discuss and the answer would be tied to the angle of the research interest, sought by enacting such a methods assemblage.

Recently, more conceptual work on data and datafication has been published attending both to empirical, theoretical, and methodological aspects of data studies, broadly defined, and adjucent fields (see e.g. Beaulieu & Leonelli, 2021; K. Braun & Kropp, 2021; Elliott, 2022; Hepp et al., 2022; Kitchin, 2021a, 2021b; Lindgren, 2020; Ruppert & Scheel, 2021; Schuilenburg & Peeters, 2021). My thesis speaks to and contributes to this literature and to the emerging, interdisciplinary field of data studies. Methods assemblages constructed in my thesis present a way to move beyond the established methodological binary 'cuts' between qualitative-quantitative, computational-'traditional', single method-mixed methods and acknowledge situated positions researchers hold in datafied societies and in academia. With the developed heuristic, I approach the goal I set for my thesis in exploring conceptually and methodologically how knowledge about datafication is co-produced with the methods assemblages. I also discussed what kinds of sensitivities can be forwarded through such a methodological reflection in data studies. The categories such as degree of collectivity of actors involved in datafication processes, their positionings in these processes, the means through which these processes are enacted, and kinds of knowledges according to which methods assemblages gather various elements and practices to a cohesive research endeavour, provide such sensitivity. Addressing data studies as an interdisciplinary body of work, my thesis also contributes to the manifold of other research domains in which empirical research on datafication is conducted, notably that of media and cultural studies and communication research building one of the backbones of data studies alongside with data science. Being able to learn from a number of leading datafication scholars in different academic disciplines and fields both through a literature

study and through expert interviews, I also explored possible developments and trends for the emerging field of data studies. Here, these trends are discussed through the lens of care-ful critique.

With my conceptual and empirical analysis of methods assemblages applied for studying datafication processes empirically, I make several contributions to the field of data studies and adjacent lines of research. *First*, I map out methods assemblages enacted in empirical research on datafication. These methods assemblages and a heuristic for their analysis together can be understood as a reflection tool. They provide a set of categories for reflecting on empirical data studies that further sensitivities to ontological and epistemological multiplicities of datafication processes and its research. These methods assemblages turn our attention away from methodological debates in studies of datafication and draw it to the empirical phenomena in which datafication research is situated. *Second*, I argued that by enacting these methods assemblages, new, re-situated conceptualisations of datafication processes are being produced. Reflecting on these re-situated conceptualisations with the methods assemblages and the heuristic I developed allows backtracking how academic knowledge production in data studies is organised. Thus, in addition to the contributions it makes to data studies, my thesis also speaks to the literature in philosophy and sociology of science. *Third*, the mapping of these re-situated conceptualisations of datafication processes in itself furthers our understanding of what the term 'datafication' stands for empirically. *Fourth*, with my methodological and conceptual inquiry into empirical research on datafication I also attend to one of the core questions of data studies: the relation between digital data, society, and knowledge. Specifically, by developing methods assemblages I explored how in data studies, various societal processes and practices are tied together with datafication processes in an endeavour to advance academic knowledge. *Fifth*, I discuss the role of critique in data studies, mapping existing critical discourses and sketching pathways for new ones, forwarded with the help of a reflection tool I developed in my thesis. *Finally*, I join the line of literature bridging data studies with feminist traditions of thought, particularly with feminist ethics of care by proposing the notion of care-ful, generative critique in data studies. It extends not only to researchers of datafied societies working in academia, but also to the practitioners in other societal domains, concerned with and interested in the ongoing datafication processes in our social reality. Care-ful and generative critique in data studies can be understood as one of the ways forward to go beyond identifying the stakes of datafication and rather in engaging locally with the manifold of relations continuously assembled and re-assembled in the processes which I addressed here with the term 'datafication'.

Future research could also expand the work I advanced with my thesis by extending the sample to such research fields as algorithms studies, data science, digital humanities, information science, and studies of digital archives. Drawing on such broad sample, future research could engage in developing further methods assemblages that encompass the technological foundation of datafication processes more directly. Moreover, deeper and more extensive engagement with the notion and practice of critique in data studies is required, if we are to take further steps to closing the 'impasse' between academic knowledge in data studies and capitalism mentioned by one of the interviewed experts (I21) in the beginning of chapter 7. With the advancement of our understandings of datafication processes, the pathways for conducting more generative and care-ful data studies would also need refinement. The pathways I sketched here speak to a broad range of empirical studies, addressed in my simple and beyond: for conducting care-ful data studies practically, more detailed pathways need to be traced that address not the field in general, but any particular local methods assemblage. I believe, in taking performativity of methods in data studies as an analytical point of departure, I could further our understanding and practices of engaging with complexity of our datafied societies in a productive way.

REFERENCES

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Acker, A., & Donovan, J. (2019). Data craft: A theory/methods package for critical internet studies. *Information, Communication & Society*, *22*(11), 1590–1609. https://doi.org/10.1080/1369118X.2019.1645194

Adam, E. (2016, January 18). Deutsche fürchten Big Data. *Süddeutsche Zeitung*. https://www.sueddeutsche.de/wirtschaft/studie-deutsche-fuerchten-big-data-1.2822871

Adams, R. (2020, July 8). Nearly 40% of A-level result predictions to be downgraded in England. *The Guardian*. https://www.theguardian.com/education/2020/aug/07/a-level-result-predictions-to-be-downgraded-england

Addey, C., & Piattoeva, N. (2021a). What a mess: Intimacies, metaphysics, multiple senses and matters of concern in education policy research (an introduction). In C. Addey & N. Piattoeva (Eds.), *Intimate Accounts of Education Policy Research* (pp. 1–15). Routledge.

Addey, C., & Piattoeva, N. (Eds.). (2021b). *Intimate Accounts of Education Policy Research: The Practice of Methods*. Routledge. https://doi.org/10.4324/9781003123613

Agostinho, D., D'Ignazio, C., Ring, A., Thylstrup, N. B., & Veel, K. (2019). Uncertain Archives: Approaching the Unknowns, Errors, and Vulnerabilities of Big Data through Cultural Theories of the Archive. *Surveillance & Society*, *17*(3/4), 422–441. https://doi.org/10.24908/ss.v17i3/4.12330

Airoldi, M. (2021). Digital traces of taste: Methodological pathways for consumer research. *Consumption Markets & Culture*, *24*(1), 97–117. https://doi.org/10.1080/10253866.2019.1690998

Alaimo, C., & Kallinikos, J. (2017). Computing the everyday: Social media as data platforms. *The Information Society*, *33*(4), 175–191. https://doi.org/10.1080/01972243.2017.1318327

Alaimo, C., & Kallinikos, J. (2021). Managing by Data: Algorithmic Categories and Organizing. *Organization Studies*, *42*(9), 1385–1407. https://doi.org/10.1177/0170840620934062

Alexander, V. D., Blank, G., & Hale, S. A. (2018). Digital traces of distinction? Popular orientation and user-engagement with status hierarchies in TripAdvisor reviews of cultural organizations. *New Media & Society*, *20*(11), 4218–4236. https://doi.org/10.1177/1461444818769448

Amoore, L. (2011). Data Derivatives: On the Emergence of a Security Risk Calculus for Our Times. *Theory, Culture & Society*, *28*(6), 24–43. https://doi.org/10.1177/0263276411417430

Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. https://www.wired.com/2008/06/pb-theory/

Apple. (n.d.). *Privacy-Preserving Contact Tracing—Apple and Google*. Apple. Retrieved 28 February 2022, from https://www.apple.com/covid19/contacttracing

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Arora, P. (2016). Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication*, *10*(0), 19.

Ashmore, M. (2015). Reflexivity in Science and Technology Studies. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 93–97). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.85018-7

Aula, V. (2019). Institutions, infrastructures, and data friction – Reforming secondary use of health data in Finland. *Big Data & Society*, *6*(2), 2053951719875980. https://doi.org/10.1177/2053951719875980

Back, L., & Puwar, N. (2012). A manifesto for live methods: Provocations and capacities. *The Sociological Review*, *60*(S1), 6–17. https://doi.org/10.1111/j.1467-954X.2012.02114.x

Baker, K. S., & Karasti, H. (2018). Data care and its politics: Designing for local collective data management as a neglected thing. *Proceedings of the 15th Participatory Design Conference on Full Papers - PDC '18*, 1–12. https://doi.org/10.1145/3210586.3210587

Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*, *28*(3), 801–831.

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Barry Born, T. (2021). Proactive state geographies: Geocoded intelligence in London's 'suburban shanty towns'. *Environment and Planning D: Society and Space*, *39*(4), 609–626. https://doi.org/10.1177/02637758211030925

Bates, J. (2018). The politics of data friction. *Journal of Documentation*, *74*(2), 412–429. https://doi.org/10.1108/JD-05-2017-0080

Bates, J. (2022). Big data and data analytics. In A. Elliott (Ed.), *The Routledge social science handbook of AI* (pp. 281–294). Routledge.

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, *3*(2), 2053951716654502. https://doi.org/10.1177/2053951716654502

Baygi, R. M., Introna, L. D., & Hultin, L. (2021). Everything Flows: Studying Continuous Socio-Technological Transformation in a Fluid and Dynamic Digital World. *MIS Quarterly*, *45*(1), 423–452. https://doi.org/10.25300/MISQ/2021/15887

Beaulieu, A., & Leonelli, S. (2021). *Data and Society. A critical introduction* (1st ed.). Sage Publications Ltd. https://pure.rug.nl/ws/portalfiles/portal/194707067/85257620_7115722_Data_and_Society_Preprint_1.pdf

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, *6*(1), 2053951718819569. https://doi.org/10.1177/2053951718819569

Beer, D. (2016). How should we do the history of Big Data? *Big Data & Society*, *3*(1), 2053951716646135. https://doi.org/10.1177/2053951716646135

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Beer, D. (2018). Envisioning the power of data analytics. *Information, Communication & Society*, *21*(3), 465–479. https://doi.org/10.1080/1369118X.2017.1289232

Beer, D., & Burrows, R. (2013). Popular Culture, Digital Archives and the New Social Life of Data. *Theory, Culture & Society*, *30*(4), 47–71. https://doi.org/10.1177/0263276413476542

Beetham, H., Drumm, L., Bell, F., Mycroft, L., & Forsythe, G. (2022). Curation and collaboration as activism: Emerging critical practices of #FemEdTech. *Learning, Media and Technology*, *47*(1), 143–155. https://doi.org/10.1080/17439884.2021.2018607

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.

Bennett, C. L., & Rosner, D. K. (2019). The Promise of Empathy: Design, Disability, and Knowing the 'Other'. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300528

Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, *12*.

Berry, D. M. (Ed.). (2012). *Understanding digital humanities*. Palgrave Macmillan. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=1339737728*

Berry, D. M. (2015 [2011]). *The Philosophy of Software—Code and Mediation in the Digital Age*. Palgrave Macmillan. https://www.palgrave.com/gp/book/9780230244184

Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of Machine Learning Research 81*, *81*, 1–11. https://papers.ssrn.com/abstract=3086546

Birchall, C. (2016). Shareveillance: Subjectivity between open and closed data. *Big Data & Society*, *3*(2), 2053951716663965. https://doi.org/10.1177/2053951716663965

Birhane, A. (2020). Algorithmic Colonization of Africa. *SCRIPTed: A Journal of Law, Technology and Society*, *17*(2), 389–409.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Birhane, A., & Guest, O. (2021). Towards Decolonising Computational Sciences. *Kvinder, Køn & Forskning*, *1*, 60–73. https://doi.org/10.7146/kkf.v29i2.124899

Blackman, L. (2019). Haunted Data, Transmedial Storytelling, Affectivity: Attending to 'Controversies' as Matters of Ghostly Concern. *Ephemera: Theory and Politics in Organisation*, *19*(1), 31–52.

Bleeker, M., Verhoeff, N., & Werning, S. (2020). Sensing data: Encountering data sonifications, materializations, and interactives as knowledge objects. *Convergence*, *26*(5–6), 1088–1107. https://doi.org/10.1177/1354856520938601

Blöbaum, B., Nölleke, D., & Scheu, A. M. (2016). Das Experteninterview in der Kommunikationswissenschaft. In S. Averbeck-Lietz & M. Meyen (Eds.), *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft* (pp. 175–190). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-01656-2_11

Boellstorff, T. (2013). Making big data, in theory. *First Monday*, *18*(10). https://doi.org/10.5210/fm.v18i10.4869

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). *Ethnography and Virtual Worlds: A Handbook of Method*. Princeton University Press.

Bogle, A. (2018, March 14). Can we trust big data to build our giant future cities? *ABC News*. https://www.abc.net.au/news/2018-03-15/big-data-algorithms-equality-future-cities/9544138

Bogner, A., & Menz, W. (2009). The Theory-Generating Expert Interview: Epistemological Interest, Forms of Knowledge, Interaction. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 43–80). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_3

Bolin, G. (2022). The Value Dynamics of Data Capitalism: Cultural Production and Consumption in a Datafied World. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *New Perspectives in Critical Data Studies*. Palgrave Macmillan Cham. https://link.springer.com/book/9783030961794

Bolin, G., & Andersson Schwarz, J. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society*, *2*(2), 2053951715608406. https://doi.org/10.1177/2053951715608406

Bonde Thylstrup, N., Flyverbom, M., & Helles, R. (2019). Datafied knowledge production: Introduction to the special theme. *Big Data & Society*, *6*(2), 2053951719875985. https://doi.org/10.1177/2053951719875985

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press.

Borgman, C. L. (2019). The Lives and After Lives of Data. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.9a36bdb6

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, *14*(6), 19–47. https://doi.org/10.1177/053901847501400602

Bowe, E., Simmons, E., & Mattern, S. (2020). Learning from lines: Critical COVID data visualizations and the quarantine quotidian. *Big Data & Society*, *7*(2), 2053951720939236. https://doi.org/10.1177/2053951720939236

Bowker, G. C. (2005). *Memory practices in the sciences*. MIT Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press. https://www.ics.uci.edu/~gbowker/classification/

boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Bradbury, A., & Roberts-Holmes, G. (2017). *The Datafication of Primary and Early Years Education: Playing with Numbers*. Routledge. https://doi.org/10.4324/9781315279053

Braidotti, R. (2019). *Posthuman knowledge*. Polity.

Braun, K., & Kropp, C. (Eds.). (2021). *In digitaler Gesellschaft: Neukonfigurationen zwischen Robotern, Algorithmen und Usern*. transcript. https://www.transcript-verlag.de/978-3-8376-5453-0/in-digitaler-gesellschaft/?c=311023049&number=978-3-8394-5453-4

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Breiter, A. (1995). Die Forschung über Künstliche Intelligenz und ihre sanduhrförmige Entwicklungsdynamik. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *2*(47), 295–318.

Breiter, A., & Hepp, A. (2018). The Complexity of Datafication: Putting Digital Traces in Context. In A. Hepp, A. Breiter, & U. Hasebrink (Eds.), *Communicative Figurations: Transforming Communications in Times of Deep Mediatization* (pp. 387–405). Springer International Publishing. https://doi.org/10.1007/978-3-319-65584-0_16

Bridges, L. E. (2021). Digital failure: Unbecoming the "good" data subject through entropic, fugitive, and queer data. *Big Data & Society*, *8*(1), 2053951720977882. https://doi.org/10.1177/2053951720977882

Brown, J. S., & Duguid, P. (2000). *The Social Life of Information*. Harvard Business School Press.

Bruns, A., & Burgess, J. (2016). Methodological Innovation in Precarious Spaces: The Case of Twitter. In H. Snee, C. Hine, Y. Morey, S. Roberts, & H. Watson (Eds.), *Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation* (pp. 17–33). Palgrave Macmillan UK. https://doi.org/10.1057/9781137453662_2

Brunsdon, C. (2020). Modelling epidemics: Technical and critical issues in the context of COVID-19. *Dialogues in Human Geography*, *10*(2), 250–254. https://doi.org/10.1177/2043820620934328

Bryman, A., & Burgess, R. G. (Eds.). (1994). *Analyzing qualitative data*. Routledge.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media and Society*, *14*(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Bucher, T. (2016). Neither Black Nor Box: Ways of Knowing Algorithms. In S. Kubitschko & A. Kaun (Eds.), *Innovative Methods in Media and Communication Research* (pp. 81–98). Springer International Publishing. https://doi.org/10.1007/978-3-319-40700-5_5

Bucher, T. (2018). *If...Then: Algorithmic Power and Politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190493028.001.0001

Burgess, M. (2020, August 20). The lessons we all must learn from the A-levels algorithm debacle. *Wired UK*. https://www.wired.co.uk/article/gcse-results-alevels-algorithm-explained

Burrows, R., & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1), 2053951714540280. https://doi.org/10.1177/2053951714540280

Candea, M. (2013). The Fieldsite as Device. *Journal of Cultural Economy*, *6*(3), 241–258. https://doi.org/10.1080/17530350.2012.754366

Carrozza, C., & Pereira, T. S. (2015). Mapping digital methods: Where science and technology studies and communication studies meet? *Communicating and Evaluating Science*, 203–236.

Chan, N. K., & Humphreys, L. (2018). Mediatization of Social Space and the Case of Uber Drivers. *Media and Communication*, *6*(2), 29–38. https://doi.org/10.17645/mac.v6i2.1316

Chang, E. (2018). Digital Meritocracy: Intermediary Organizations and the Construction of Policy Knowledge. *Educational Policy*, 0895904818802116. https://doi.org/10.1177/0895904818802116

Cheney-Lippold, J. (2017). *We Are Data: Algorithms and The Making of Our Digital Selves*. NYU Press. https://doi.org/10.2307/j.ctt1gk0941

Christin, A. (2020). *Metrics at Work*. Princeton University Press. https://press.princeton.edu/books/ebook/9780691200002/metrics-at-work

Chun, W. H. K. (2005). On Software, or the Persistence of Visual Knowledge. *Grey Room*, *18*, 26–51. https://doi.org/10.1162/1526381043320741

Chun, W. H. K. (2011). *Programmed Visions: Software and Memory*. MIT Press.

Cifor, M., Garcia, P., Cowan, T. L., Rault, J., Sutherland, T., Chan, A., Rode, J., Hoffmann, A. L., Salehi, N., & Nakamura, L. (2019). *Feminist Data Manifest-No*. Feminist Data Manifest-No. https://www.manifestno.com

Cinnamon, J. (2019). *Visual Data Justice? Datafication of Urban Informality in South Africa Using 360° Imaging Technologies* (SSRN Scholarly Paper ID 3456196). Social Science Research Network. https://doi.org/10.2139/ssrn.3456196

Cinnamon, J. (2020). Attack the Data: Agency, Power, and Technopolitics in South African Data Activism. *Annals of the American Association of Geographers*, *110*(3), 623–639. https://doi.org/10.1080/24694452.2019.1644991

Cochoy, F., Hagberg, J., McIntyre, M. P., & Sörum, N. (2017). *Digitalizing Consumption: How devices shape consumer culture*. Taylor & Francis.

Cohen, B. R., & Galusky, W. (2010). Guest Editorial. *Science as Culture*, *19*(1), 1–14. https://doi.org/10.1080/09505430903557924

Coleman, R. (2018). Theorizing the present: Digital media, pre-emergence and infra-structures of feeling. *Cultural Studies*, *32*(4), 600–622. https://doi.org/10.1080/09502386.2017.1413121

Coleman, R., Page, T., & Palmer, H. (2019, May 15). Feminist New Materialist Practice: The Mattering of Methods. *MAI: Feminism & Visual Culture*. https://maifeminism.com/feminist-new-materialisms-the-mattering-of-methods-editors-note/

Coletta, C., & Kitchin, R. (2017). Algorhythmic governance: Regulating the 'heartbeat' of a city using the Internet of Things. *Big Data & Society*, *4*(2), 2053951717742418. https://doi.org/10.1177/2053951717742418

Collins, H. M., & Pinch, T. (2012 [1998]). *The Golem: What You Should Know About Science* (2nd, 13th printing ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107295612

Coopmans, C., & McNamara, K. M. (2020). Care in Translation: Care-ful Research in Medical Settings. *East Asian Science, Technology and Society: An International Journal*, *14*(1), 1–14. https://doi.org/10.1215/18752160-8193729

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press. https://design-justice.pubpub.org/pub/ap8rgw5e/release/1

Couldry, N. (2004). Theorising media as practice. *Social Semiotics*, *14*(2), 115–132. https://doi.org/10.1080/1035033042000238295

Couldry, N. (2012). *Media, Society, World: Social Theory and Digital Media Practice*. Polity.

Couldry, N. (2020). Recovering critique in an age of datafication. *New Media & Society*, *22*(7), 1135–1151. https://doi.org/10.1177/1461444820912536

Couldry, N., & Hepp, A. (2017). *The mediated construction of reality*. Polity Press.

Couldry, N., Macdonald, R., Stephansen, H. C., Clark, W., Dickens, L., & Fotopoulou, A. (2015). Constructing a digital storycircle: Digital infrastructure and mutual recognition. *International Journal of Cultural Studies*, *18*, 501–517.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Couldry, N., & Yu, J. (2018). Deconstructing datafication's brave new world. *New Media & Society*, *20*(12), 4473–4491. https://doi.org/10.1177/1461444818775968

Craig, R. T. (1999). Communication Theory as a Field. *Communication Theory*, *9*(2), 119–161. https://doi.org/10.1111/j.1468-2885.1999.tb00355.x

Craig, R. T. (2018). For a Practical Discipline. *Journal of Communication*, *68*(2), 289–297. https://doi.org/10.1093/joc/jqx013

Crawford, K., Gray, M. L., & Miltner, K. (2014). Big Data| Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction. *International Journal of Communication*, *8*(0), 10.

Crawford, K., & Joler, V. (2018). *"Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources"*. AI Now Institute and Share Lab. https://anatomyof.ai

Currie, M. E., Paris, B. S., & Donovan, J. M. (2019). What difference do data make? Data management and social change. *Online Information Review*, *43*(6), 971–985. https://doi.org/10.1108/OIR-02-2018-0052

Czarniawska, B. (2007). *Shadowing: And Other Techniques for Doing Fieldwork in Modern Societies*. Copenhagen Business School Press DK.

Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, *3*(1), 205395171664834. https://doi.org/10.1177/2053951716648346

Dalton, C. M., & Thatcher, J. (2014). *What Does A Critical Data Studies Look Like, And Why Do We Care?* Space and Society. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care

Decuypere, M. (2019). Researching educational apps: Ecologies, technologies, subjectivities and learning regimes. *Learning, Media and Technology*, *0*(0), 1–16. https://doi.org/10.1080/17439884.2019.1667824

Decuypere, M. (2021). The Topologies of Data Practices: A Methodological Introduction. *Journal of New Approaches in Educational Research*, *10*(1), 67–84. https://doi.org/10.7821/naer.2021.1.650

Decuypere, M., & Broeck, P. V. (2020). Time and educational (re-)forms—Inquiring the temporal dimension of education. *Educational Philosophy and Theory*, *0*(0), 1–11. https://doi.org/10.1080/00131857.2020.1716449

Delfanti, A. (2021). Machinic dispossession and augmented despotism: Digital work in an Amazon warehouse. *New Media & Society*, *23*(1), 39–55. https://doi.org/10.1177/1461444819891613

Dencik, L. (2019). Situating practices in datafication – from above and below. In H. C. Stephansen & E. Treré (Eds.), *Citizen Media and Practice* (pp. 243–255). Routledge.

Dencik, L. (2020). Mobilizing Media Studies in an Age of Datafication. *Television & New Media*, *21*(6), 568–573. https://doi.org/10.1177/1527476420918848

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring Data Justice: Conceptions, Applications and Directions. *Information, Communication & Society*, *22*(7), 873–881. https://doi.org/10.1080/1369118X.2019.1606268

Dencik, L., & Kaun, A. (2020). Datafication and the Welfare State. *Global Perspectives*, *1*(1). https://doi.org/10.1525/gp.2020.12912

Dencik, L., & Stevens, S. (2021). Regimes of justification in the datafied workplace: The case of hiring. *New Media & Society*, 14614448211052892. https://doi.org/10.1177/14614448211052893

Denzin, N. K., & Lincoln, Y. S. (2018). Introduction: The Discipline amd practice of Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (Fifth edition, pp. 1–26). SAGE.

Der AMS-Algorithmus. (n.d.). https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus

Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Diaz-Bone, R., Horvath, K., & Cappel, V. (2020). Social Research in Times of Big Data: The Challenges of New Data Worlds and the Need for a Sociology of Social Research. *Historical Social Research*, *45*(3), 314–341. https://doi.org/10.12759/hsr.45.2020.3.314-341

Didier, E. (2022). Free from Numbers? The Politics of Qualitative Sociology in the U.S. Since 1945. In A. Mennicken & R. Salais (Eds.), *The New Politics of Numbers* (pp. 417–463). Springer International Publishing. https://doi.org/10.1007/978-3-030-78201-6_13

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press.

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Computational Communication Science | Outlining the Way Ahead in Computational Communication Science: An Introduction to the IJoC Special Section on "Computational Methods for Communication Science: Toward a Strategic Roadmap". *International Journal of Communication*, *13*(0), 9.

Donia, J., & Shaw, J. A. (2021). Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. *Big Data & Society*, *8*(2), 20539517211065250. https://doi.org/10.1177/20539517211065248

Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society*, *3*(2), 2053951716665128. https://doi.org/10.1177/2053951716665128

Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, *5*(2), 2053951718784083. https://doi.org/10.1177/2053951718784083

Dowd, C. (2016). The new order of news and social media enterprises: Visualisations, linked data, and new methods and practices in journalism. *Communication Research and Practice*, *2*(1), 97–110. https://doi.org/10.1080/22041451.2016.1155339

Draude, C., Klumbyte, G., Lücking, P., & Treusch, P. (2019). Situated algorithms: A sociotechnical systemic approach to bias. *Online Information Review*, *44*(2), 325–342. https://doi.org/10.1108/OIR-10-2018-0332

Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral Vision: Expertise in Real World Contexts. *Organization Studies*, *26*(5), 779–792. https://doi.org/10.1177/0170840605053102

Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, *005*(1).

Elliott, A. (2022). *The Routledge social science handbook of AI*. Routledge.

Estalella, A. (2016). Prototyping Social Sciences: Emplacing Digital Methods. In H. Snee, C. Hine, Y. Morey, S. Roberts, & H. Watson (Eds.), *Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation* (pp. 127–142). Palgrave Macmillan UK. https://doi.org/10.1057/9781137453662_8

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Edition). St. Martin's Press.

European Commission. (n.d.). *Coronavirus: EU interoperability gateway* [Text]. European Commission. Retrieved 28 February 2022, from https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_1905

Feyerabend, P. (2010 [1988]). *Against Method*. Verso.

Fiesler, C., Young, A., Peyton, T., Bruckman, A. S., Gray, M., Hancock, J., & Lutters, W. (2015). Ethics for Studying Online Sociotechnical Systems in a Big Data World. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, 289–292. https://doi.org/10.1145/2685553.2685558

Fileborn, B., & Trott, V. (2021). "It ain't a compliment": Feminist data visualisation and digital street harassment advocacy. *Convergence*, 13548565211045536. https://doi.org/10.1177/13548565211045536

Fink, A. (2005). *Conducting research literature reviews: From the Internet to paper* (2nd ed). Sage Publications. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=88802265*

Fiore-Gartland, B., & Neff, G. (2015). Communication, Mediation, and the Expectations of Data: Data Valences Across Health and Wellness Communities. *International Journal of Communication*, *9*(0), 19.

Fleck, L. (2019 [1935]). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv* (L. Schäfer & T. Schnelle, Eds.; 12. Auflage). Suhrkamp.

Flensburg, S., & Lomborg, S. (2021). Datafication research: Mapping the field for a future agenda. *New Media & Society*, 14614448211046616. https://doi.org/10.1177/14614448211046616

Flick, U. (2011). *Triangulation: Eine Einführung* (3rd ed.). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92864-7

Fligstein, N., & McAdam, D. (2012). *A Theory of Fields*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199859948.001.0001

Flyverbom, M., Leonardi, P., Stohl, C., & Stohl, M. (2016). Digital Age| The Management of Visibilities in the Digital Age—Introduction. *International Journal of Communication*, *10*(0), 12.

Flyverbom, M., & Murray, J. (2018). Datastructuring—Organizing and curating digital traces into action. *Big Data & Society*, *5*(2), 2053951718799114. https://doi.org/10.1177/2053951718799114

Forlano, L. (2016). Hacking the Feminist Disabled Body » The Journal of Peer Production. *Journal of Peer Production*, *8*. http://peerproduction.net/issues/issue-8-feminism-and-unhacking-2/peer-reviewed-papers/issue-8-feminism-and-unhackingpeer-reviewed-papers-2hacking-the-feminist-disabled-body/

Forlano, L. (2017). Data Rituals in Intimate Infrastructures: Crip Time and the Disabled Cyborg Body as an Epistemic Site of Feminist Science | Catalyst: Feminism, Theory, Technoscience. *Catalyst: Feminism, Theory, Technoscience*, *3*(2), 1–28. https://doi.org/10.28968/cftt.v3i2.28843

Fotopoulou, A. (2019). Understanding citizen data practices from a feminist perspective: Embodiment and the ethics of care. In *Citizen Media and Practice*. Routledge.

Fotopoulou, A. (2020). Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information, Communication & Society*, *0*(0), 1–18. https://doi.org/10.1080/1369118X.2020.1716041

Fourcade, M. (2007). The Politics of Method and Its Agentic, Performative, and Ontological Others. *Social Science History*, *31*(1), 107–114. https://doi.org/10.1017/S0145553200013675

Fox, N. J., & Alldred, P. (2018). Mixed methods, materialism and the micropolitics of the research-assemblage. *International Journal of Social Research Methodology*, *21*(2), 191–204. https://doi.org/10.1080/13645579.2017.1350015

Freeman, R. (2009). What is 'translation'? *Evidence & Policy: A Journal of Research, Debate and Practice*, *5*(4), 429–447. https://doi.org/10.1332/174426409X478770

Fuchs, C. (2019). What is Critical Digital Social Research? Five Reflections on the Study of Digital Society. *Journal of Digital Social Research*, *1*(1), 10–16. https://doi.org/10.33621/jdsr.v1i1.7

Fuller, S. (2012). CSI: Kuhn and Latour. *Social Studies of Science*, *42*(3), 429–434. https://doi.org/10.1177/0306312712437228

Galloway, A. R. (2006). Language Wants To Be Overlooked: On Software and Ideology. *Journal of Visual Culture*, *5*(3), 315–331. https://doi.org/10.1177/1470412906070519

Garcia, S. G., & Barclay, K. (n.d.). *Adapting research methodologies in the COVID-19 pandemic. Resources for researchers* (pp. 1–24). e Nippon Foundation Ocean Nexus at EarthLab, University of Washingto. https://www.uts.edu.au/sites/default/files/2020-06/uts-adapting-research-methodologies-covid-19-pandemic-resources-researchers.pdf

Gardner, P., & Kember, S. (2021). Introduction: Probing the System: Feminist Complications of Automated Technologies, Flows, and Practices of Everyday Life. *Catalyst: Feminism, Theory, Technoscience*, *7*(2), Article 2. https://doi.org/10.28968/cftt.v7i2.36962

Garforth, L. (2012). In/Visibilities of Research: Seeing and Knowing in STS. *Science, Technology, & Human Values*, *37*(2), 264–285. https://doi.org/10.1177/0162243911409248

Gentzel, P. (2015). *Praxistheorie und Mediatisierung*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-08994-8

Gentzel, P. (2019). Materialität, Technik und das Subjekt: Elemente kritischer Kommunikations- und Medienanalyse. In P. Gentzel, F. Krotz, J. Wimmer, & R. Winter (Eds.), *Das vergessene Subjekt: Subjektkonstitutionen in mediatisierten Alltagswelten* (pp. 87–113). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23936-7_5

Gentzel, P., Krotz, F., Wimmer, J., & Winter, R. (2019). *Das Vergessene Subjekt: Subjektkonstitutionen in Mediatisierten Alltagswelten*. Vieweg. https://link.springer.com/book/10.1007/978-3-658-23936-7

Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, *15*(8), 1348–1365. https://doi.org/10.1177/1461444812472322

Gerlitz, C., Helmond, A., Nieborg, D. B., & van der Vlist, F. N. (2019). Apps and Infrastructures – a Research Agenda. *Computational Culture*, *7*. http://computationalculture.net/apps-and-infrastructures-a-research-agenda/

Gerrard, Y., & Bates, J. (2019). Introduction to the Data Power Special Issue: Tactics, access and shaping. *Online Information Review*, *43*(6), 945–951. https://doi.org/10.1108/OIR-10-2019-396

Gherardi. (2019). *How to Conduct a Practice-based Study*. Edward Elgar Publishing. https://www.elgaronline.com/view/9781788973557/9781788973557.xml

Gillespie, T. (2010). The politics of 'platforms'. *New Media and Society*, *12*(3), 347–364. https://doi.org/10.1177/1461444809342738

Gitelman, L. (2013). *Raw Data Is an Oxymoron*. MIT Press.

Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen* (4. Auflage). VS Verlag. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=1697302440*

Glatt, Z. (2021). *LSE Digital Ethnography Collective Reading List SHARED DOC*. LSE Digital Ethnography Collective. https://zoeglatt.com/wp-content/uploads/2021/05/LSE-Digital-Ethnography-Collective-Reading-List-SHARED-DOC-May-2021.pdf

Google. (n.d.). *Use the COVID-19 Exposure Notifications System on your Android phone—Android Help*. Retrieved 28 February 2022, from https://support.google.com/android/answer/9888358?hl=en

Gorur, R. (2018). Escaping Numbers? Intimate Accounting and the challenge to numbers in Australia's 'Education Revolution'. *Science & Technology Studies*, *31*(4), 89–108. https://doi.org/10.23987/sts.56745

Gorur, R., & Dey, J. (2021). Making the user friendly: The ontological politics of digital data platforms. *Critical Studies in Education*, *62*(1), 67–81. https://doi.org/10.1080/17508487.2020.1727544

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 2053951718786316. https://doi.org/10.1177/2053951718786316

Gray, J., & Witt, A. (2021). A feminist data ethics of care for machine learning: The what, why, who and how. *First Monday*. https://doi.org/10.5210/fm.v26i12.11833

Greene, J. C. (2006). Toward a Methodology of Mixed Methods Social Inquiry. *Research in the Schools*, *13*(1), 93–98.

Gregory, K., McMillan Cottom, T., & Daniels, J. (2016). Introduction. In J. Daniels, K. Gregory, & T. McMillan Cottom (Eds.), *Digital Sociologies*. Policy Press.

Greiffenhagen, C., Mair, M., & Sharrock, W. (2011). From Methodology to Methodography: A Study of Qualitative and Quantitative Reasoning in Practice. *Methodological Innovations Online*, *6*(3), 93–107. https://doi.org/10.4256/mio.2011.009

Grek, S., Maroy, C., & Verger, A. (Eds.). (2020). *World Yearbook of Education 2021: Accountability and Datafication in the Governance of Education*. Routledge. https://doi.org/10.4324/9781003014164

Grenz, T., & Kirschner, H. (2018). Digital Traces in Context| Unraveling the App Store: Toward an Interpretative Perspective on Tracing. *International Journal of Communication*, *12*(0), 17.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Halkier, B. (2017). Questioning the 'Gold Standard' Thinking in Qualitative Methods from a Practice Theoretical Perspective: Towards Methodological Multiplicity. In M. Jonas, B. Littig, & A. Wroblewski (Eds.), *Methodological Reflections on Practice Oriented Theories* (pp. 193–204). Springer International Publishing. https://doi.org/10.1007/978-3-319-52897-7_13

Hand, M., & Hillyard, S. (2014). *Big data?: Qualitative approaches to digital research*. Emerald. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=908919

Haraway, D. J. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599. JSTOR. https://doi.org/10.2307/3178066

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.

Hargittai, E. (Ed.). (2020). *Research Exposed: How Empirical Social Science Gets Done in the Digital Age* (p. 288 Pages). Columbia University Press. https://suche.suub.uni-bremen.de/peid=B174432789&LAN=DE&CID=6700502&index=L&Hitnr=1&dtyp=O&rtyp=a

Hart, C. (1998). *Doing a literature review: Releasing the research imagination*. Sage. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=3504997834*

Hartong, S., & Piattoeva, N. (2021). Contextualizing the datafication of schooling – a comparative discussion of Germany and Russia. *Critical Studies in Education*, *62*(2), 227–242. https://doi.org/10.1080/17508487.2019.1618887

Hasinoff, A., & Bivens, R. (2021). Feature Analysis: A Method for Analyzing the Role of Ideology in App Design. *Journal of Digital Social Research*, *3*(2), 89-113-89–113. https://doi.org/10.33621/jdsr.v3i2.56

Heeks, R. (2018). *Information and Communication Technology for Development (ICT4D)*. Routledge.

Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean? *Information Development*, *34*(1), 90–102. https://doi.org/10.1177/0266666916678282

Helles, R., & Flyverbom, M. (2019). Meshes of Surveillance, Prediction, and Infrastructure: On the Cultural and Commercial Consequences of Digital Platforms. *Surveillance & Society*, *17*(1/2), 34–39. https://doi.org/10.24908/ss.v17i1/2.13120

Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society*, *1*(2), 2056305115603080. https://doi.org/10.1177/2056305115603080

Hepp, A. (2016). Pioneer communities: Collective actors in deep mediatisation. *Media, Culture & Society*, *38*(6), 918–933. https://doi.org/10.1177/0163443716664484

Hepp, A. (2020). *Deep mediatization*. Routledge.

Hepp, A., Breiter, A., & Friemel, T. N. (2018). Digital Traces in Context| Digital Traces in Context—An Introduction. *International Journal of Communication*, *12*, 11.

Hepp, A., Jarke, J., & Kramp, L. (Eds.). (2022). *New Perspectives in Critical Data Studies*. Palgrave Macmillan Cham. https://link.springer.com/book/9783030961794

Hepp, A., Loosen, W., & Hasebrink, U. (2021). Jenseits des Computational Turn: Methodenentwicklung und Forschungssoftware in der Kommunikations- und Medienwissenschaft – zur Einführung in das Themenheft. *M&K Medien & Kommunikationswissenschaft*, *69*(1), 3–24. https://doi.org/10.5771/1615-634X-2021-1-3-1

Heuer, H. (2020). *Users & Machine Learning-based Curation Systems* [Universität Bremen]. https://media.suub.uni-bremen.de/handle/elib/4444

Heuer, H., Jarke, J., & Breiter, A. (2021). Machine learning in tutorials – Universal applicability, underinformed application, and other misconceptions. *Big Data & Society*, *8*(1), 20539517211017590. https://doi.org/10.1177/20539517211017593

Hirschauer, S. (2008). Die Empiriegeladenheit von Theorien und der Erfindungsreichtum der Praxis. In S. Hirschauer, H. Kalthoff, & G. Lindemann (Eds.), *Theoretische Empirie. Zur Relevanz qualitativer Forschung* (pp. 165–187). Suhrkamp.

Hobart, H. J. K., & Kneese, T. (2020). Radical Care: Survival Strategies for Uncertain Times. *Social Text*, *38*(1 (142)), 1–16. https://doi.org/10.1215/01642472-7971067

Hockenhull, M., & Cohn, M. L. (2021). Speculative Data Work & Dashboards: Designing Alternative Data Visions. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 264:1-264:31. https://doi.org/10.1145/3434173

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hui, A., Schatzki, T., Shove, E., Schatzki, T., & Shove, E. (2016). *The Nexus of Practices: Connections, constellations, practitioners*. Routledge. https://doi.org/10.4324/9781315560816

Hultin, L. (2019). On becoming a sociomaterial researcher: Exploring epistemological practices grounded in a relational, performative ontology. *Information and Organization*, *29*(2), 91–104. https://doi.org/10.1016/j.infoandorg.2019.04.004

Hurdley, R., & Dicks, B. (2011). In-between practice: Working in the 'thirdspace' of sensory and multimodal methodology. *Qualitative Research*, *11*(3), 277–292. https://doi.org/10.1177/1468794111399837

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 2053951716674238. https://doi.org/10.1177/2053951716674238

Introna, L. D. (2015). Algorithms, Governance, and Governmentality: On Governing Academic Writing. *Science, Technology, & Human Values*, *41*(1), 17–49. https://doi.org/10.1177/0162243915587360

Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society*, *16*(3), 169–185. https://doi.org/10.1080/01972240050133634

Isin, E., & Ruppert, E. (2019). Data's empire: Postcolonial data politics. In D. Bigo, E. Isin, & E. Ruppert (Eds.), *Data Politics*. Routledge.

Jarke, J. (2019). Open government for all? Co-creating digital public services for older adults through data walks. *Online Information Review*. https://doi.org/10.1108/OIR-02-2018-0059

Jarke, J. (2021). *Co-creating Digital Public Services for an Ageing Society: Evidence for User-centric Design*. Springer International Publishing. https://doi.org/10.1007/978-3-030-52873-7

Jarke, J., & Breiter, A. (2019). Editorial: The datafication of education. *Learning, Media and Technology*, *44*(1), 1–6. https://doi.org/10.1080/17439884.2019.1573833

Jarke, J., & Zakharova, I. (forthcoming). Researching the datafication of organisation through assemblage- and process-oriented data journeys. *Historical Social Research / Historische Sozialforschung*. https://www.gesis.org/hsr/

Jensen, K. B., & Neumann, W. R. (2013). Evolving Paradigms of Communication Research. *International Journal of Communication*, *7*, 230–238.

Jonas, M., Littig, B., & Wroblewski, A. (2017). Object, Perspectives and Methodology of Praxeological Research. In M. Jonas, B. Littig, & A. Wroblewski (Eds.), *Methodological Reflections on Practice Oriented Theories* (pp. 251–261). Springer International Publishing. https://doi.org/10.1007/978-3-319-52897-7_17

Jones, K. M. L., & McCoy, C. (2019). Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework. *Learning, Media and Technology*, *44*(1), 52–63. https://doi.org/10.1080/17439884.2018.1556216

Kara, H. (2020). *Creative Research Methods 2e: A Practical Guide* (2. Edition). Policy Press.

Karlsson, A. (2019). A Room of One's Own?: Using period trackers to escape menstrual stigma. *Nordicom Review*, *40*(s1), 111–123. https://doi.org/10.2478/nor-2019-0017

Katzenbach, C., Magalhães, J., Kopps, A., Sühr, T., & Wunderlich, L. (2021). *The Platform Governance Archive (PGA)*. https://doi.org/10.17605/OSF.IO/XSBPT

Kaun, A. (2021). Suing the algorithm: The mundanization of automated decision-making in public services through litigation. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118X.2021.1924827

Kaun, A., & Stiernstedt, F. (2020). Prison media work: From manual labor to the work of being tracked. *Media, Culture & Society*, *42*(7–8), 1277–1292. https://doi.org/10.1177/0163443719899809

Kaun, A., Velkova, J., Laaksonen, S.-M., Delfanti, A., Logsdon, A., Stiernstedt, F., Lehtiniemi, T., & Ruckenstein, M. (2020). IN THE SHADOWS OF THE DIGITAL ECONOMY: THE GHOST WORK OF INFRASTRUCTURAL LABOR. *AoIR Selected Papers of Internet Research*. https://doi.org/10.5210/spir.v2020i0.11131

Kennedy, H. (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.

Kennedy, H. (2019). Book Review: Noortje Marres, Digital Sociology: The Re-Invention of Social Research. *Sociology*, *53*(1), 216–218. https://doi.org/10.1177/0038038517732257

Kennedy, H., & Bates, J. (2017). Data Power in Material Contexts: Introduction. *Television & New Media*, *18*(8), 701–705. https://doi.org/10.1177/1527476417720034

Kennedy, H., & Hill, R. L. (2018). The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation. *Sociology*, *52*(4), 830–848. https://doi.org/10.1177/0038038516674675

Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society*, *2*(2), 2053951715611145. https://doi.org/10.1177/2053951715611145

Kennedy, H., Moss, G., Birchall, C., & Moshonas, S. (2015). Balancing the potential and problems of digital methods through action research: Methodological reflections. *Information, Communication & Society*, *18*(2), 172–186. https://doi.org/10.1080/1369118X.2014.946434

Kennedy, H., Oman, S., Taylor, M., Bates, J., & Steedman, R. (2020). *Public understanding and perceptions of data practices: A review of existing research*. University of Sheffield. http://livingwithdata.org/current-research/publications/

Kennedy, H., Steedman, R., & Jones, R. (2021). Approaching public perceptions of datafication through the lens of inequality: A case study in public service media. *Information, Communication & Society*, *24*(12), 1745–1761. https://doi.org/10.1080/1369118X.2020.1736122

King, N., & Brooks, J. M. (2016). *Template Analysis for Business and Management Students*. SAGE.

Kitchin, R. (2014a). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kitchin, R. (2014b). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014c). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE. https://doi.org/10.4135/9781473909472

Kitchin, R. (2021a). *Data Lives: How Data Are Made and Shape our World*. Bristol University Press. https://doi.org/10.46692/9781529215649

Kitchin, R. (2021b). *The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures*. SAGE.

Kitchin, R., & Lauriault, T. (2014). *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work* (SSRN Scholarly Paper ID 2474112). Social Science Research Network. https://papers.ssrn.com/abstract=2474112

Kjær, K. M., Ojala, M., & Henriksen, L. (2021). Absent Data: Engagements with Absence in a Twitter Collection Process. *Catalyst: Feminism, Theory, Technoscience*, *7*(2), Article 2. https://doi.org/10.28968/cftt.v7i2.34563

Klumbytė, G. (2022). Algorithmic Kinning. *Matter: Journal of New Materialist Research*, *3*(1), Article 1. https://doi.org/10.1344/jnmr.v3i1.38962

Knorr-Cetina, K. (2002). *Wissenskulturen: Ein Vergleich naturwissenschaftlicher Wissensformen*. Suhrkamp.

Knorr-Cetina, K. (2016 [1984]). *Die Fabrikation von Erkenntnis: Zur Anthropologie der Naturwissenschaft* (Erweiterte Neuauflage, 4. Auflage). Suhrkamp.

Koble, N. (2019, June 6). The complicated truth about China's social credit system. *Wired UK*. https://www.wired.co.uk/article/china-social-credit-system-explained

Koch, G. (2019). *Zur „Datafication" der Wissens- produktion in der qualitativen Forschung: Vol. Forschungsdesign 4.0-Datengenerierung und Wissenstransfer in interdisziplinärer Perspektive* (E. Bünz, W. Müller, J. Schneider, & I. Spieker, Eds.). Institut für Sächsische Geschichte und Volkskunde. https://slub.qucosa.de/api/qucosa%3A35929/attachment/ATT-0/

Koenen, E., Schwarzenegger, C., & Kittler, J. (2021). Data(fication): "Understanding the World Through Data" as an Everlasting Revolution. In *Digital Roots* (pp. 137–156). De Gruyter Oldenbourg. https://doi.org/10.1515/9783110740202-008

Konopásek, Z. (2008). Making Thinking Visible with Atlas.ti: Computer Assisted Qualitative Analysis as Textual Practices. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *9*(2), Article 2. https://doi.org/10.17169/fqs-9.2.420

Koopman, C. (2019). *How We Became Our Data: A Genealogy of the Informational Person*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/H/bo38181810.html

Koro-Ljungberg, M., & Mazzei, L. A. (2012). Problematizing Methodological Simplicity in Qualitative Research: Editors' Introduction. *Qualitative Inquiry*, *18*(9), 728–731. https://doi.org/10.1177/1077800412453013

Kotliar, D. M. (2020). The return of the social: Algorithmic identity in an age of symbolic demise. *New Media & Society*, *22*(7), 1152–1167. https://doi.org/10.1177/1461444820912535

Kozinets, R. V. (2010). *Netnography: Ethnographic research in the age of the internet*. Sage Publications Ltd.

Krasmann, S. (2020). The logic of the surface: On the epistemology of algorithms in times of big data. *Information, Communication & Society*, *23*(14), 2096–2109. https://doi.org/10.1080/1369118X.2020.1726986

Krippendorff, K. (1970). On Generating Data in Communication Research. *Journal of Communication*, *20*(3), 241–269. https://doi.org/10.1111/j.1460-2466.1970.tb00883.x

Krippendorff, K. (2016). Data. In *The International Encyclopedia of Communication Theory and Philosophy* (pp. 1–6). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118766804.wbiect104

Kubicek, H., Breiter, A., & Jarke, J. (2019). Daten, Metadaten, Interoperabilität. In T. Klenk, F. Nullmeier, & G. Wewer (Eds.), *Handbuch Digitalisierung in Staat und Verwaltung* (pp. 1–13). Springer Fachmedien. https://doi.org/10.1007/978-3-658-23669-4_1-1

Kubitschko, S. (2018). Acting on media technologies and infrastructures: Expanding the media as practice approach. *Media, Culture & Society*, *40*(4), 629–635. https://doi.org/10.1177/0163443717706068

Kubitschko, S., & Kaun, A. (2016). *Innovative Methods in Media and Communication Research*. Springer.

Kuch, D., Kearnes, M., & Gulson, K. (2020). The promise of precision: Datafication in medicine, agriculture and education. *Policy Studies*, *41*(5), 527–546. https://doi.org/10.1080/01442872.2020.1724384

Kuhn, T. S. (2020 [1976]). *Die Struktur wissenschaftlicher Revolutionen* (H. Vetter, Trans.; Zweite revidierte und um das Postskriptum von 1969 ergänzte Auflage, [26. Auflage]). Suhrkamp.

Lai, S. S., & Flensburg, S. (2020a). A proxy for privacy uncovering the surveillance ecology of mobile apps. *Big Data & Society*, *7*(2), 2053951720942543. https://doi.org/10.1177/2053951720942543

Lai, S. S., & Flensburg, S. (2020b). Appscapes in everyday life: Studying Mobile Datafication from an Infrastructural User Perspective. *MedieKultur: Journal of Media and Communication Research*, *36*(69), 029–051. https://doi.org/10.7146/mediekultur.v36i69.121018

Langley, A., & Ravasi, D. (2019). Visual Artefacts as Tools for Analysis and Theorizing. In T. B. Zilber, J. M. Amis, & J. Mair (Eds.), *Research in the Sociology of Organizations* (Vol. 59, pp. 173–199). Emerald Publishing Limited. https://doi.org/10.1108/S0733-558X20190000059010

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, *30*(2), 225–248.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press.

Laudon, K. C., & Laudon, J. P. (2020). *Management information systems: Managing the digital firm* (Sixteenth edition). Pearson.

Lauriault, T. P., & Lim, M. (2019). Special Section on Data Power. *Canadian Journal of Communication*, *44*(3), Article 3. https://doi.org/10.22230/cjc.2019v44n3a3647

Lauriault, T. P., & Taylor, D. R. F. (2019). Chapter 11—The preservation and archiving of geospatial data and Cybercartography as a proactive preservation process. In D. R. F. Taylor, E. Anonby, & K. Murasugi (Eds.), *Modern Cartography Series* (Vol. 7, pp. 179–196). Academic Press. https://doi.org/10.1016/B978-0-444-64193-9.00011-7

Law, J. (2004). *After method mess in social science research*. Routledge ; Netlibrary [distributor. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=115106

Law, J. (2021). From After Method to care-ful research (a foreword). In C. Addey & N. Piattoeva (Eds.), *Intimate Accounts of Education Policy Research* (pp. xvi–xx). Routledge.

Law, J., & Lin, W. (2020). *Care-ful Research: Sensibilities from STS*. 17.

Law, J., & Ruppert, E. (2013). THE SOCIAL LIFE OF METHODS: Devices. *Journal of Cultural Economy*, *6*(3), 229–240. https://doi.org/10.1080/17530350.2013.812042

Law, J., Ruppert, E., & Savage, M. (2011). *The double social life of methods. Centre for Research on Socio-Cultural Change (CRESC), Working Paper No 95*.

Law, J., & Singleton, V. (2000). Performing Technology's Stories: On Social Constructivism, Performance, and Performativity. *Technology and Culture*, *41*(4), 765–775. https://doi.org/10.1353/tech.2000.0167

Law, J., & Singleton, V. (2013). ANT and Politics: Working in and on the World. *Qualitative Sociology*, *36*(4), 485–502. https://doi.org/10.1007/s11133-013-9263-7

Law, J., & Urry, J. (2004). Enacting the social. *Economy and Society*, *33*(3), 390–410. https://doi.org/10.1080/0308514042000225716

Lazer, D., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lazer, D., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170

Leavitt, C. M. (2019). Keeping Software Present: Software as a Timely Object for STS Studies of the Digital. In *DigitalSTS: A Field Guide for Science & Technology Studies*. Princeton University Press. https://doi.org/10.1515/9780691190600-029

Lehtiniemi, T., & Haapoja, J. (2020). Data agency at stake: MyData activism and alternative frames of equal participation. *New Media & Society*, *22*(1), 87–104. https://doi.org/10.1177/1461444819861955

Leonardi, P. M., & Treem, J. W. (2020). Behavioral Visibility: A new paradigm for organization studies in the age of digitization, digitalization, and datafication. *Organization Studies*, *41*(12), 1601–1625. https://doi.org/10.1177/0170840620970728

Leonelli, S. (2019). Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.17405bb6

Leonelli, S. (2020). Learning from Data Journeys. In S. Leonelli & N. Tempini (Eds.), *Data Journeys in the Sciences* (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7_1

Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Big Data & Society*, *8*(1), 20539517211010310. https://doi.org/10.1177/20539517211010310

Leonelli, S., Rappert, B., & Davies, G. (2017). Data Shadows: Knowledge, Openness, and Absence. *Science, Technology, & Human Values*, *42*(2), 191–202. https://doi.org/10.1177/0162243916687039

Lewis, S., & Holloway, J. (2019). Datafying the teaching 'profession': Remaking the professional teacher in the image of data. *Cambridge Journal of Education*, *49*(1), 35–51. https://doi.org/10.1080/0305764X.2018.1441373

Lewis, T., Gangadharan, S. P., Saba, M., & Petty, T. (2018). *Digital Defense Playbook: Community Power Tools for Reclaiming Data*. Our Data Bodies. https://www.odbproject.org/wp-content/uploads/2019/03/ODB_DDP_HighRes_Single.pdf

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Lindell, J. (2020). Establishment versus Newcomers, Critical versus Administrative?: Sketching the structure of the Swedish field of media and communication studies. *Nordicom Review*, *41*(2), 109–125. https://doi.org/10.2478/nor-2020-0009

Lindén, L. (2021). Moving Evidence: Patients' Groups, Biomedical Research, and Affects. *Science, Technology, & Human Values*, *46*(4), 815–838. https://doi.org/10.1177/0162243920948126

Lindén, L., & Lydahl, D. (2021). Editorial: Care in STS. *Nordic Journal of Science and Technology Studies*, 3–12. https://doi.org/10.5324/njsts.v9i1.4000

Lindgren, S. (2019). Hacking Social Science for the Age of Datafication. *Journal of Digital Social Research*, *1*(1), 1–9. https://doi.org/10.33621/jdsr.v1i1.6

Lindgren, S. (2020). *Data theory: Interpretive sociology and computational methods*. Polity. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=3770647270*

Lingard, B. (2011). Policy as numbers: Ac/counting for educational research. *The Australian Educational Researcher*, *38*(4), 355–382. https://doi.org/10.1007/s13384-011-0041-9

Lippert, I., & Verran, H. (2018). After Numbers? Innovations in Science and Technology Studies' Analytics of Numbers and Numbering. *Science & Technology Studies*, 2–12. https://doi.org/10.23987/sts.76416

Livingstone, S. (2019). Audiences in an Age of Datafication: Critical Questions for Media Research. *Television & New Media*, *20*(2), 170–183. https://doi.org/10.1177/1527476418811118

Livingstone, S., & Sefton-Green, J. (2016). *The class: Living and learning in the digital age*. New York University Press.

Lohr, S. (2012, February 11). The Age of Big Data. *The New York Times*. https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Lomborg, S., Dencik, L., & Moe, H. (2020). Methods for datafication, datafication of methods: Introduction to the Special Issue. *European Journal of Communication*, 0267323120922045. https://doi.org/10.1177/0267323120922045

Lomborg, S., Langstrup, H., & Andersen, T. O. (2020). Interpretation as luxury: Heart patients living with data doubt, hope, and anxiety. *Big Data & Society*, *7*(1), 2053951720924436. https://doi.org/10.1177/2053951720924436

López-Gómez, D. (2019). What if ANT wouldn't pursue agnosticism but care? In A. Blok, I. Farias, & C. Roberts (Eds.), *The Routledge Companion to Actor-Network Theory* (pp. 4–13). Routledge.

Lu, A. J., Dillahunt, T. R., Marcu, G., & Ackerman, M. S. (2021). Data Work in Education: Enacting and Negotiating Care and Control in Teachers' Use of Data-Driven Classroom

Surveillance Technology. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 452:1-452:26. https://doi.org/10.1145/3479596

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

Lupton, D. (2015). *Digital sociology*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315776880

Lupton, D. (2016). *The Quantified Self. A Sociology of Self-Tracking*. Polity.

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

Lupton, D. (2020). Thinking With Care About Personal Data Profiling: A More-Than-Human Approach. *International Journal of Communication*, *14*(0), 3165–3183.

Lupton, D. (2021). *DOING FIELDWORK IN A PANDEMIC*. https://docs.google.com/document/d/1clGjGABB2h2qbduTgfqribHmog9B6P0NvMgVuiHZCl8/edit?usp=embed_facebook

Lupton, D., & Watson, and A. (2020). Towards more-than-human digital data studies: Developing research-creation methods. *Qualitative Research*, 1468794120939235. https://doi.org/10.1177/1468794120939235

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, *19*(5), 780–794. https://doi.org/10.1177/1461444816686328

Lury, C. (2020). *Problem spaces: How and why methodology matters*. Polity Press.

Lury, C., Tironi, M., & Bernasconi, R. (2020). The Social Life of Methods as Epistemic Objects: Interview with Celia Lury. *Diseña*, *16*, 32–55. https://doi.org/10.7764/disena.16.32-55

Lury, C., & Wakeford, N. (Eds.). (2012). *Inventive methods: The happening of the social*. Routledge.

Macgilchrist, F. (2021). Theories of Postdigital Heterogeneity: Implications for Research on Education and Datafication. *Postdigital Science and Education*, *3*(3), 660–667. https://doi.org/10.1007/s42438-021-00232-w

Mackenzie, A., & Vurdubakis, T. (2011). Codes and Codings in Crisis: Signification, Performativity and Excess. *Theory, Culture & Society*, *28*(6), 3–23. https://doi.org/10.1177/0263276411424761

MacKenzie, D. A., & Wajcman, J. (Eds.). (1999). *The social shaping of technology* (2nd ed). Open University Press.

Madianou, M. (2019). Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises. *Social Media + Society*, *5*(3), 2056305119863146. https://doi.org/10.1177/2056305119863146

Madsen, A. K. (2018). Data in the smart city: How incongruent frames challenge the transition from ideal to practice. *Big Data & Society*, *5*(2), 2053951718802321. https://doi.org/10.1177/2053951718802321

Manning, E. (2015). Against method. In *Non-representational methodologies* (pp. 62–81). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315883540-8/method-erin-manning

Manolev, J., Sullivan, A., & Slee, R. (2019). The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Learning, Media and Technology*, *44*(1), 36–51. https://doi.org/10.1080/17439884.2018.1558237

Manovich, L. (2013). *Software takes command: Extending the language of new media*. Bloomsbury.

Markham, A. N. (2013). Undermining 'data': A critical examination of a core term in scientific inquiry. *First Monday*, *18*(10). https://doi.org/10.5210/fm.v18i10.4868

Markham, A. N. (2019). Critical Pedagogy as a Response to Datafication. *Qualitative Inquiry*, *25*(8), 754–760. https://doi.org/10.1177/1077800418809470

Marres, N. (2012). The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived. *The Sociological Review*, *60*(1_suppl), 139–165. https://doi.org/10.1111/j.1467-954X.2012.02121.x

Marres, N. (2017). *Digital sociology: The reinvention of social research*. Polity.

Marres, N., & Gerlitz, C. (2016). Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology: *The Sociological Review*. https://journals.sagepub.com/doi/10.1111/1467-954X.12314

Mascheroni, G., & Siibak, A. (2021). *Datafied Childhoods*. Peter Lang Verlag. https://www.peterlang.com/document/1140627

Masson, E. (2017). Humanistic Data Research: An Encounter between Epistemic Traditions. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society: Studying Culture through Data* (pp. 25–37). Amsterdam University Press. https://doi.org/10.25969/mediarep/12400

Mathieu, D., & Jorge, A. (2020). The Datafication of Media (and) Audiences: An Introduction. *MedieKultur: Journal of Media and Communication Research*, *36*(69), 001–010. https://doi.org/10.7146/mediekultur.v36i69.122585

Mathieu, D., & Pruulmann Vengerfeldt, P. (2020). The Data loop of media and audience: How audiences and media actors make datafication work. *MedieKultur: Journal of Media and Communication Research*, *36*(69), 116–138. https://doi.org/10.7146/mediekultur.v36i69.121178

Mattern, S. (2018). Maintenance and Care. *Places Journal*. https://doi.org/10.22269/181120

Mattoni, A. (2017). A situated understanding of digital technologies in social movements. Media ecology and media practice approaches. *Social Movement Studies*, *16*(4), 494–505. https://doi.org/10.1080/14742837.2017.1311250

Mattoni, A. (2018). From Data Extraction to Data Leaking. Data-activism in Italian and Spanish anti-corruption campaigns. *PARTECIPAZIONE E CONFLITTO*, *10*(3), 723-746–746. https://doi.org/10.1285/i20356609v10i3p723

Mau, S. (2017). *Das metrische Wir: Über die Quantifizierung des Sozialen* (Erste Auflage, Originalausgabe). Suhrkamp.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. Murray.

Mcglotten, S. (2016). Black Data. In P. E. Johnson (Ed.), *No Tea, No Shade* (pp. 262–286). Duke University Press. https://doi.org/10.1515/9780822373711-016

McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, *7*(1), 2053951720904386. https://doi.org/10.1177/2053951720904386

Medina Perea, I. A. (2021). *Socio-material factors shaping patient data journeys in the United Kingdom*. University of Sheffield.

Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, *8*(4), Article 4. https://policyreview.info/

Metz, C., & Satariano, A. (2020, February 6). An Algorithm That Grants Freedom, or Takes It Away. *The New York Times*. https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2), 2053951720948087. https://doi.org/10.1177/2053951720948087

Micheli, M., Tsinaraki, C., Minghini, M., Mitton, I., Schade, S., & Kotsev, A. (2022). A European approach to COVID-19 apps? A landscape analysis of apps stores in the initial stages of the health crisis. *First Monday*. https://doi.org/10.5210/fm.v27i3.12412

Milan, S. (2019). Acting on Data(fication). In H. C. Stephansen & E. Treré (Eds.), *Citizen Media and Practice: Currents, Connections, Challenges* (pp. 212–226). Routledge. https://papers.ssrn.com/abstract=3489068

Milan, S., & Beraldo, D. (2019). From Data Politics to the Contentious Politics of Data. *Big Data and Society*, *6*(2). https://doi.org/10.1177/2053951719885967

Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television & New Media*, *20*(4), 319–335. https://doi.org/10.1177/1527476419837739

Milan, S., & Velden, L. van der. (2016). The Alternative Epistemologies of Data Activism. *Digital Culture & Society*, *2*(2), 57–74. https://doi.org/10.14361/dcs-2016-0205

Miller, P., & Rose, N. (1990). Governing economic life. *Economy and Society*, *19*(1), 1–31. https://doi.org/10.1080/03085149000000001

Milmo, D., & Farah, H. (2022, March 5). War as seen on TikTok: Ukraine clips get views whether true or not. *The Guardian*. https://www.theguardian.com/technology/2022/mar/05/tiktok-ukraine-russia-invasion-clips-get-views-whether-true-or-not

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press.

Mol, A. (2008). *The logic of care: Health and the problem of patient choice*. Routledge. http://site.ebrary.com/lib/suub/detail.action?docID=10236177

Mol, A. (2010). Actor-Network Theory: Sensitive terms and enduring tensions. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie. Sonderheft*, *50*. https://dare.uva.nl/search?identifier=75bbc661-0a89-475a-9eef-8c8c5a2e9904

Mol, A., Moser, I., & Pols, J. (2010). *Care in Practice: On Tinkering in Clinics, Homes and Farms*. transcript. https://doi.org/10.14361/transcript.9783839414477

Moor, L., & Uprichard, E. (2014). The Materiality of Method: The Case of the Mass Observation Archive. *Sociological Research Online*, *19*(3), 10.

Mountz, A., Bonds, A., Mansfield, B., Loyd, J., Hyndman, J., Walton-Roberts, M., Basu, R., Whitson, R., Hawkins, R., Hamilton, T., & Curran, W. (2015). For Slow Scholarship: A Feminist Politics of Resistance through Collective Action in the Neoliberal University. *ACME: An International Journal for Critical Geographies*, *14*(4), 1235–1259.

Mützel, S. (2015). Facing Big Data: Making sociology relevant. *Big Data & Society*, *2*(2), 2053951715599179. https://doi.org/10.1177/2053951715599179

n.a. (2021, July 21). Investigation: How TikTok's Algorithm Figures Out Your Deepest Desires. *Wall Street Journal*. https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how-tiktok-algorithm-figures-out-your-deepest-desires/6C0C2040-FF25-4827-8528-2BD6612E3796

Nake, F. (2002). Data, Information, and Knowledge. In K. Liu, R. J. Clarke, P. B. Andersen, R. K. Stamper, & E.-S. Abou-Zeid (Eds.), *Organizational Semiotics: Evolving a Science of Information Systems* (pp. 41–50). Springer US. https://doi.org/10.1007/978-0-387-35611-2_3

NASA. (n.d.). *Climate Change Evidence: How Do We Know?* Climate Change: Vital Signs of the Planet. Retrieved 8 March 2022, from https://climate.nasa.gov/evidence

Neff, G., Tanweer, A., Fiore-Gartland, B., & Osburn, L. (2017). Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science. *Big Data*, *5*(2), 85–97. https://doi.org/10.1089/big.2016.0050

Neumayer, C., Rossi, L., & Struthers, D. M. (2021). Invisible Data: A Framework for Understanding Visibility Processes in Social Media Data. *Social Media + Society*, *7*(1), 2056305120984472. https://doi.org/10.1177/2056305120984472

Nicolini, D. (2009a). Articulating Practice through the Interview to the Double. *Management Learning*, *40*(2), 195–212. https://doi.org/10.1177/1350507608101230

Nicolini, D. (2009b). Zooming In and Out: Studying Practices by Switching Theoretical Lenses and Trailing Connections. *Organization Studies*, *30*(12), 1391–1418. https://doi.org/10.1177/0170840609349875

Nicolini, D. (2017). Practice Theory as a Package of Theory, Method and Vocabulary: Affordances and Limitations. In M. Jonas, B. Littig, & A. Wroblewski (Eds.), *Methodological Reflections on Practice Oriented Theories* (pp. 19–34). Springer International Publishing. https://doi.org/10.1007/978-3-319-52897-7_2

Nikunen, K. (2021). Ghosts of white methods? The challenges of Big Data research in exploring racism in digital context. *Big Data & Society*, *8*(2), 20539517211048964. https://doi.org/10.1177/20539517211048964

Nind, M., Wiles, R., Bengry-Howell, A., & Crow, G. (2013). Methodological innovation and research ethics: Forces in tension or forces in harmony? *Qualitative Research*, *13*(6), 650–667. https://doi.org/10.1177/1468794112455042

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1497317

Nordfalk, F. (2022). The Mutual Enablement of Research Data and Care: How Newborn Babies Become a National Research Population. *Science & Technology Studies*, *35*(1), 21–34. https://doi.org/10.23987/sts.98655

Nordstrom, S. N. (2015). Not So Innocent Anymore: Making Recording Devices Matter in Qualitative Interviews. *Qualitative Inquiry*, *21*(4), 388–401. https://doi.org/10.1177/1077800414563804

Nordstrom, S. N. (2018). Antimethodology: Postqualitative Generative Conventions. *Qualitative Inquiry*, *24*(3), 215–226. https://doi.org/10.1177/1077800417704469

Onwuegbuzie, A. J., & Frels, R. (2016). *Seven steps to a comprehensive literature review: A multimodal and cultural approach*. Sage.

Onwuegbuzie, A. J., & Leech, N. L. (2005). On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. *International Journal of Social Research Methodology*, *8*(5), 375–387. https://doi.org/10.1080/13645570500402447

Onwuegbuzie, A. J., Leech, N. L., & Collins, K. M. T. (2011). Innovative Qualitative Data Collection Techniques for Conducting Literature Reviews/Research Syntheses. In M. Williams & W. Vogt, *The SAGE Handbook of Innovation in Social Research Methods* (pp. 182–204). SAGE Publications Ltd. https://doi.org/10.4135/9781446268261.n13

Özkula, S. M., Reilly, P. J., & Hayes, J. (2022). Easy data, same old platforms? A systematic review of digital activism methodologies. *Information, Communication & Society*, *0*(0), 1–20. https://doi.org/10.1080/1369118X.2021.2013918

Pangrazio, L. (2020, October 30). New ways of understanding digital data. *Materialising Data*. https://materialising-data.org/2020/10/30/metaphors-of-data-its-time-to-redefine-how-we-talk-about-data/

Papacharissi, Z. (Ed.). (2018). *A networked self and platforms, stories, connections*. Routledge, Taylor & Francis Group.

Paßmann, J., & Boersma, A. (2017). Unknowing Algorithms. On Transparency of Unopenable Black Boxes. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 139–146). Amsterdam University Press. https://doi.org/10.5117/9789462981362

Passoth, J.-H., & Wehner, J. (Eds.). (2013). *Quoten, Kurven und Profile*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-93139-5

Paul, K., & Anguiano, D. (2021, October 23). Facebook crisis grows as new whistleblower and leaked documents emerge. *The Guardian*. https://www.theguardian.com/technology/2021/oct/22/facebook-whistleblower-hate-speech-illegal-report

Pellegrino, G., Söderberg, J., & Milan, S. (2019). Datafication from Below: Epistemology, Ambivalences, Challenges. *TECNOSCIENZA: Italian Journal of Science & Technology Studies*, *10*(1), 89–114.

Peña Gangadharan, S., & Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society*, *22*(7), 882–899. https://doi.org/10.1080/1369118X.2019.1593484

Pentzold, C. (2015). Praxistheoretische Prinzipien, Traditionen und Perspektiven kulturalistischer Kommunikations- und Medienforschung. *Medien & Kommunikationswissenschaft*, *63*(2), 229–245. https://doi.org/10.5771/1615-634x-2015-2-229

Perez, C. C. (2019). *Invisible Women: Exposing Data Bias in a World Designed for Men* (1. Edition). Chatto & Windus.

Perrotta, C., Selwyn, N., & Ewin, C. (2022). Artificial intelligence and the affective labour of understanding: The intimate moderation of a language model. *New Media & Society*, 14614448221075296. https://doi.org/10.1177/14614448221075296

Perrotta, C., & Williamson, B. (2018). The social life of Learning Analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, *43*(1), 3–16. https://doi.org/10.1080/17439884.2016.1182927

Piattoeva, N., & Saari, A. (2020). Rubbing against data infrastructure(s): Methodological explorations on working with(in) the impossibility of exteriority. *Journal of Education Policy*, *0*(0), 1–21. https://doi.org/10.1080/02680939.2020.1753814

Pickering, A. (2012). The world since Kuhn. *Social Studies of Science*, *42*(3), 467–473. https://doi.org/10.1177/0306312712440503

Pinch, T. (2015). Scientific Controversies. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 281–286). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.85043-6

Pinch, T., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Pinel, C., Prainsack, B., & McKevitt, C. (2020). Caring for data: Value creation in a data-intensive research laboratory. *Social Studies of Science*, *50*(2), 175–197. https://doi.org/10.1177/0306312720906567

Pinel, C., & Svendsen, M. N. (2021). In search of 'extra data': Making tissues flow from personal to personalised medicine. *Big Data & Society*, *8*(2), 20539517211035664. https://doi.org/10.1177/20539517211035664

Pink, S. (2012). *Situating Everyday Life: Practices and Places*. SAGE.

Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, *5*(1), 205395171775322. https://doi.org/10.1177/2053951717753228

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press. https://muse.jhu.edu/book/77331

Posada, J. (2022). Embedded reproduction in platform data work. *Information, Communication & Society*, *0*(0), 1–19. https://doi.org/10.1080/1369118X.2022.2049849

Posner, M., & Klein, L. F. (2017). Editor's IntroductionData as Media. *Feminist Media Histories*, *3*(3), 1–8. https://doi.org/10.1525/fmh.2017.3.3.1

Postill, J. (2010). Introduction: Theorising Media and Practice. In B. Bräuchler & J. Postill (Eds.), *Theorising media and practice* (pp. 1–34). Berghahn Books. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=1171549873*

Prietl, B. (2019). The Politics of Datafication. *Komplexe Dynamiken globaler und lokaler Entwicklungen. Verhandlungen des 39. Kongresses der Deutschen Gesellschaft für Soziologie in Göttingen 2018.*, *39*. https://publikationen.soziologie.de/index.php/kongressband_2018/article/view/1060

Prietl, B., & Houben, D. (2018). Einführung. Soziologische Perspektiven auf die Datafizierung der Gesellschaft. In D. Houben & B. Prietl (Eds.), *Datengesellschaft* (pp. 7–32). transcript Verlag. https://doi.org/10.1515/9783839439579-001

Prinsloo, P. (2019). A social cartography of analytics in education as performative politics. *British Journal of Educational Technology*, *50*(6), 2810–2823. https://doi.org/10.1111/bjet.12872

Pruulmann-Vengerfeldt, P., & Hörste, H. M. zu. (2020). Reimagining Audiences in the Age of Datafication. In *Reimagining Communication: Experience*. Routledge.

Puig de la Bellacasa, M. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, *41*(1), 85–106. https://doi.org/10.1177/0306312710380301

Puig de la Bellacasa, M. (2017). *Matters of Care: Speculative Ethics in More than Human Worlds*. University of Minnesota Press. https://www.jstor.org/stable/10.5749/j.ctt1mmfspt

Puschmann, C., & Burgess, J. (2014). Big data, big questions| Metaphors of big data. *International Journal of Communication*, *8*, 20.

Pybus, J., & Coté, M. (2021). Did you give permission? Datafication in the mobile ecosystem. *Information, Communication & Society*, *0*(0), 1–19. https://doi.org/10.1080/1369118X.2021.1877771

Rajão, R., & Jarke, J. (2018). The materiality of data transparency and the (re)configuration of environmental activism in the Brazilian Amazon. *Social Movement Studies*, *17*(3), 318–332. https://doi.org/10.1080/14742837.2018.1459297

Rajão, R., & Vurdubakis, T. (2013). On the Pragmatics of Inscription: Detecting Deforestation in the Brazilian Amazon. *Theory, Culture & Society*, *30*(4), 151–177. https://doi.org/10.1177/0263276413486203

Ratner, H., & Gad, C. (2018). Data warehousing organization: Infrastructural experimentation with educational governance: *Organization*. https://doi.org/10.1177/1350508418808233

Reckwitz, A. (2003). Basic Elements of a Theory of Social Practices. *Zeitschrift für Soziologie*, 20.

Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, *20*(4), 350–365. https://doi.org/10.1177/1527476419831640

Rieder, B., & Röhle, T. (2017). Digital Methods. From Challenges to Bildung. In M. T. Schäfer & K. Es, van (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 109–124). Amsterdam University Press. http://en.aup.nl/books/9789462981362-the-datafied-society.html

Rinkinen, J., Jalas, M., & Shove, E. (2015). Object Relations in Accounts of Everyday Life. *Sociology*, *49*(5), 870–885. https://doi.org/10.1177/0038038515577910

Robertson, T., & Wagner, I. (2013). Ethics: Engagement, representation and politics-in-action. In J. Simonsen & T. Robertson (Eds.), *Routledge International Handbook of Participatory Design* (pp. 64–85). Routledge.

Rogers, R. (2009). *The End of the Virtual: Digital Methods*. Amsterdam University Press; nlebk. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=305755&site=ehost-live

Rogers, R. (2013). *Digital Methods*. MIT Press. https://doi.org/10.7551/mitpress/8718.001.0001

Rogers, R. (2015). Digital Methods for Web Research. *Emerging Trends in the Social and Behavioral Sciences*. https://doi.org/10.1002/9781118900772.etrds0076

Rogers, R. (2019). *Doing digital methods* (1st edition). SAGE Publications.

Rosiek, J. L., Snyder, J., & Pratt, S. L. (2020). The New Materialisms and Indigenous Theories of Non-Human Agency: Making the Case for Respectful Anti-Colonial Engagement. *Qualitative Inquiry*, *26*(3–4), 331–346. https://doi.org/10.1177/1077800419830135

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

Ruckenstein, M., & Schüll, N. D. (2017). The Datafication of Health. *Annual Review of Anthropology*, *46*(1), 261–278. https://doi.org/10.1146/annurev-anthro-102116-041244

Ruppert, E. (2013). Rethinking empirical social sciences. *Dialogues in Human Geography*, *3*(3), 268–273. https://doi.org/10.1177/2043820613514321

Ruppert, E., Isin, E., & Bigo, D. (2017). Data politics. *Big Data & Society*, *4*(2), 2053951717717749. https://doi.org/10.1177/2053951717717749

Ruppert, E., Law, J., & Savage, M. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, *30*(4), 22–46. https://doi.org/10.1177/0263276413484941

Ruppert, E., & Scheel, S. (Eds.). (2021). *Data Practices: Making Up a European People*. Goldsmiths Press.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 2053951718820549. https://doi.org/10.1177/2053951718820549

Saifer, A., & Dacin, M. T. (2021). Data and Organization Studies: Aesthetics, emotions, discourse and our everyday encounters with data. *Organization Studies*, 01708406211006250. https://doi.org/10.1177/01708406211006250

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3E [Third edition]). SAGE. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=1595257497*

Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Salmons, J. E. (2015). *Doing Qualitative Research Online*. SAGE.

Sánchez-Monedero, J., & Dencik, L. (2019). The datafication of the workplace. *Working Paper DATA JUSTICE*, 48.

Sandberg, M., Rossi, L., Galis, V., & Bak Jørgensen, M. (Eds.). (2022). *Research Methodologies and Ethical Challenges in Digital Migration Studies: Caring For (Big) Data?* Springer Nature. https://doi.org/10.1007/978-3-030-81226-3

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, 1–23. https://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf

Savage, M. (2013). The 'Social Life of Methods': A Critical Introduction. *Theory, Culture & Society*, *30*(4), 3–21. https://doi.org/10.1177/0263276413486160

Savage, M., & Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, *41*(5), 885–899. https://doi.org/10.1177/0038038507080443

Schäfer, H. (2017). Relationality and Heterogeneity: Transitive Methodology in Practice Theory and Actor-Network Theory. In M. Jonas, B. Littig, & A. Wroblewski (Eds.), *Methodological Reflections on Practice Oriented Theories* (pp. 35–46). Springer International Publishing. https://doi.org/10.1007/978-3-319-52897-7_3

Schäfer, M. T., & van Es, K. (Eds.). (2017). *The Datafied Society. Studying Culture through Data*. Amsterdam University Press. https://doi.org/10.5117/9789462981362

Schatzki, T. R. (1996). *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge University Press.

Schatzki, T. R. (2002). *Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*. Penn State Press.

Schmidt, R. (2017). Sociology of Social Practices: Theory or Modus Operandi of Empirical Research? In M. Jonas, B. Littig, & A. Wroblewski (Eds.), *Methodological Reflections on Practice Oriented Theories* (pp. 3–17). Springer International Publishing. https://doi.org/10.1007/978-3-319-52897-7_1

Schreier, M. (2014). Ways of Doing Qualitative Content Analysis: Disentangling Terms and Terminologies. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *15*(1). https://doi.org/10.17169/fqs-15.1.2043

Schrock, A. (2017). What communication can contribute to data studies: Three lenses on communication and data. *International Journal of Communication*, *11*, 9.

Schuilenburg, M., & Peeters, R. (Eds.). (2021). *The algorithmic society: Technology, power, and knowledge*. Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780429261404

Schüll, N. D. (2016). Data for life: Wearable technology and the design of self-care. *BioSocieties*, *11*(3), 317–333. https://doi.org/10.1057/biosoc.2015.47

Schulz, A. H., & Breiter, A. (2012). Was fördert Kooperation? Betrachtung von Schulinformationssystemen durch Logfileanalysen. In J. Desel, J. M. Haake, & C. Spannagel (Eds.), *DeLFI 2012: Die 10. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.* (pp. 147–158). Gesellschaft für Informatik e.V.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 2053951717738104. https://doi.org/10.1177/2053951717738104

Sellar, S. (2015). A feel for numbers: Affect, data and education policy. *Critical Studies in Education*, *56*(1), 131–146.

Selwyn, N. (2020). 'Just playing around with Excel and pivot tables'—The realities of data-driven schooling. *Research Papers in Education*, 1–20. https://doi.org/10.1080/02671522.2020.1812107

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Sharp, G., Fardouly, J., Bromberg, M., Leaver, T., & Gerrard, Y. (2021, September 23). *Instagram can make teens feel bad about their body, but parents can help. Here's how*. The Conversation. http://theconversation.com/instagram-can-make-teens-feel-bad-about-their-body-but-parents-can-help-heres-how-168093

Shen, H., DeVos, A., Eslami, M., & Holstein, K. (2021). Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–29. https://doi.org/10.1145/3479577

Shove, E., Pantzar, M., & Watson, M. (2012). *The dynamics of social practice: Everyday life and how it changes*. SAGE.

Sismondo, S. (2010). *An introduction to science and technology studies* (2nd ed). Wiley-Blackwell. https://www.wiley.com/en-us/An+Introduction+to+Science+and+Technology+Studies%2C+2nd+Edition-p-9781405187657

Sismondo, S. (2012). Fifty years of The Structure of Scientific Revolutions, twenty-five of Science in Action. *Social Studies of Science*, *42*(3), 415–419. https://doi.org/10.1177/0306312712439255

Smith, G. J. (2018). Data doxa: The affective consequences of data practices. *Big Data & Society*, *5*(1), 2053951717751551. https://doi.org/10.1177/2053951717751551

Snee, H., Hine, C., Morey, Y., Roberts, S., & Watson, H. (2016). *Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation*. Palgrave Macmillan UK Imprint : Palgrave Macmillan. http://dx.doi.org/10.1057/9781137453662

Southerton, C. (2020). Datafication. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data* (pp. 1–4). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_332-1

Spaargaren, G., Weenink, D., & Lamers, M. (Eds.). (2016). *Practice Theory and Research: Exploring the dynamics of social life* (0 ed.). Routledge. https://doi.org/10.4324/978131565690

Springgay, S., & Truman, S. E. (2018). On the Need for Methods Beyond Proceduralism: Speculative Middles, (In)Tensions, and Response-Ability in Research. *Qualitative Inquiry*, *24*(3), 203–214. https://doi.org/10.1177/1077800417704464

St. Pierre, E. A. (2018). Writing Post Qualitative Inquiry. *Qualitative Inquiry*, *24*(9), 603–608. https://doi.org/10.1177/1077800417734567

St. Pierre, E. A. (2019). Post Qualitative Inquiry in an Ontology of Immanence. *Qualitative Inquiry*, *25*(1), 3–16. https://doi.org/10.1177/1077800418772634

Stamper, R. (1996). Signs, Information, Norms and Systems. In B. Holmqvist, P. B. Andersen, H. Klein, & R. Posner (Eds.), *Signs of Work*. DE GRUYTER. https://doi.org/10.1515/9783110819014-013

Star, S. L. (1995). *Ecologies of Knowledge: Work and Politics in Science and Technology*. SUNY Press. https://suche.suub.uni-bremen.de/peid=ebsco-6307&LAN=DE&CID=6700502&index=L&Hitnr=2&dtyp=O&rtyp=

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 253–264. https://dl.acm.org/doi/pdf/10.1145/192844.193021

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Star, S. L., & Strauss, A. (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, *8*(1), 9–30. https://doi.org/10.1023/A:1008651105359

Stark, L., & Hoffmann, A. L. (2019). Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Culture. *Journal of Cultural Analytics*, *4*(1), 11052. https://doi.org/10.22148/16.036

Steensen, S. (2019). Journalism's epistemic crisis and its solution: Disinformation, datafication and source criticism. *Journalism*, *20*(1), 185–189. https://doi.org/10.1177/1464884918809271

Stephansen, H. C. (2019). Conceptualizing the role of knowledge in acting on media. In *Citizen Media and Practice*. Routledge. https://westminsterresearch.westminster.ac.uk/download/c977364c324ac9e3fc010aa14af9bf26b50d126a07ed1466aa0039fcb7a44bcf/126297/Conceptualising%20the%20role%20of%20knowledge%20in%20%27acting%20on%20media%27_author%20accepted.pdf

Stephansen, H. C., & Treré, E. (2019). *Citizen Media and Practice: Currents, Connections, Challenges*. Routledge.

Stock, W. G., & Stock, M. (2013). Handbook of Information Science. In *Handbook of Information Science*. De Gruyter Saur. https://doi.org/10.1515/9783110235005

Strathern, M. (2000). The Tyranny of Transparency. *British Educational Research Journal*, *26*(3), 309–321. https://doi.org/10.1080/713651562

Strauss, A. L., & Corbin, J. M. (1996). *Grounded theory: Grundlagen qualitativer Sozialforschung*. Beltz, Psychologie-Verlag-Union.

Suchman, L. (2002). Located accountabilities in technology production. *Scandinavian Journal of Information Systems*, *14*(2), 16.

Suchman, L., Trigg, R., & Blomberg, J. (2002). Working artefacts: Ethnomethods of the prototype. *The British Journal of Sociology*, *53*(2), 163–179. https://doi.org/10.1080/00071310220133287

Taguchi, H. L. (2012). A diffractive and Deleuzian approach to analysing interview data. *Feminist Theory*, *13*(3), 265–281. https://doi.org/10.1177/1464700112456001

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, *4*(2), 2053951717736335. https://doi.org/10.1177/2053951717736335

Taylor, L., Sharma, G., Martin, A., & Jameson, S. (Eds.). (2020). *Data justice and Covid-19: Global perspectives*. Meatspace Press.

Thompson, T. L. (2020). Data-bodies and data activism: Presencing women in digital heritage research. *Big Data & Society*, *7*(2), 2053951720965613. https://doi.org/10.1177/2053951720965613

Thylstrup, N. B. (2019). Data out of place: Toxic traces and the politics of recycling. *Big Data & Society*, *6*(2), 2053951719875479. https://doi.org/10.1177/2053951719875479

Tkacz, N., Henrique da Mata Martins, M., Porto de Albuquerque, J., Horita, F., & Dolif Neto, G. (2021). Data diaries: A situated approach to the study of data. *Big Data & Society*, *8*(1), 2053951721996036. https://doi.org/10.1177/2053951721996036

Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge. http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=22545269*

Tronto, J. C. (2016). *Who cares?: How to reshape a democratic politics*. Cornell Selects, an imprint of Cornell University Press.

Tsai, Y.-S., Perrotta, C., & Gašević, D. (2020). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*, *45*(4), 554–567. https://doi.org/10.1080/02602938.2019.1676396

Tufekci, Z. (2014a). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Eighth International AAAI Conference on Weblogs and Social Media*.

Tufekci, Z. (2014b). Engineering the public: Big data, surveillance and computational politics. *First Monday*. https://doi.org/10.5210/fm.v19i7.4901

Turner, S. (2012). Whatever happened to knowledge? *Social Studies of Science*, *42*(3), 474–480. https://doi.org/10.1177/0306312712436555

Ulmer, J. B. (2017). Writing Slow Ontology. *Qualitative Inquiry*, *23*(3), 201–211. https://doi.org/10.1177/1077800416643994

Uprichard, E. (2013, October 1). *Focus: Big Data, Little Questions?* Discover Society. https://discoversociety.org/2013/10/01/focus-big-data-little-questions/

Vallee, M. (2020). Doing nothing does something: Embodiment and data in the COVID-19 pandemic. *Big Data & Society*, *7*(1), 2053951720933930. https://doi.org/10.1177/2053951720933930

van Atteveldt, W., margolin, drew, Shen, C., Trilling, D., & Weber, R. (2019). *A Roadmap for Computational Communication Research* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/4dhfk

van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, *12*(2–3), 81–92. https://doi.org/10.1080/19312458.2018.1458084

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

van Dijck, J. (2017). Foreword. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society. Studying Culture through Data*. Amsterdam University Press.

van Dijck, J., & Poell, T. (2013). Understanding Social Media Logic. *Media and Communication*, *1*(1), 2–24. https://doi.org/10.12924/mac2013.01010002

van Eck, N. J., & Waltman, L. (2007). VOS: A New Method for Visualizing Similarities Between Objects. In R. Decker & H.-J. Lenz (Eds.), *Advances in Data Analysis* (pp. 299–306). Springer. https://doi.org/10.1007/978-3-540-70981-7_34

van Es, K., & de Lange, M. (2020). Data with its boots on the ground: Datawalking as research method. *European Journal of Communication*, 026732312092208. https://doi.org/10.1177/0267323120922087

van Es, K., Schäfer, M. T., & Wieringa, M. (2021). Tool Criticism and the Computational Turn. A "Methodological Moment" in Media and Communication Studies. *M&K Medien & Kommunikationswissenschaft*, *69*(1), 46–64. https://doi.org/10.5771/1615-634X-2021-1-46

van Es, K., Wieringa, M., & Schäfer, M. T. (2018). Tool Criticism: From Digital Methods to Digital Methodology. *Proceedings of the 2nd International Conference on Web Studies*, 24–27. https://doi.org/10.1145/3240431.3240436

van Koningsbruggen, R., Waldschütz, H., & Hornecker, E. (2022). What is Data? - Exploring the Meaning of Data in Data Physicalisation Teaching. *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 1–21. https://doi.org/10.1145/3490149.3501319

Vannini, P. (2015). *Non-Representational Methodologies: Re-Envisioning Research*. Routledge. https://doi.org/10.4324/9781315883540

Velkova, J. (2018). *Studying Emerging Data Practices: Creating a Cultural Biography of Objects Through Using the Web as an Ethnographic Resource*. SAGE Publications Ltd. https://doi.org/10.4135/9781526445803

Velkova, J., & Kaun, A. (2019). Algorithmic resistance: Media practices and the politics of repair. *Information, Communication & Society*, *0*(0), 1–18. https://doi.org/10.1080/1369118X.2019.1657162

Venturini, T., & Latour, B. (2010). The social fabric: Digital traces and quali-quantitative methods. *Proceedings of Future En Seine*, 30–15.

Venturini, T., Latour, B., & Meunier, A. (2015). Eine unerwartete Reise: Einige Lehren über Kontinuität aus den Erfahrungen des Sciences Po médialab. In F. Süssenguth (Ed.), *Die Gesellschaft der Daten: Über die digitale Transformation der sozialen Ordnung* (pp. 17–40). Transcript. https://doi.org/10.1515/9783839427644-002

Vines, J., Clarke, R., Wright, P., McCarthy, J., & Olivier, P. (2013). Configuring Participation: On How We Involve People in Design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 429–438. https://doi.org/10.1145/2470654.2470716

Waisbord, S. R. (2019). *Communication: A post-discipline*. Polity Press.

Wanka, A., & Gallistl, V. (2018). Doing Age in a Digitized World—A Material Praxeology of Aging With Technology. *Frontiers in Sociology*, *3*. https://doi.org/10.3389/fsoc.2018.00006

Watson, A. (2020). Methods Braiding: A Technique for Arts-Based and Mixed-Methods Research. *Sociological Research Online*, *25*(1), 66–83. https://doi.org/10.1177/1360780419849437

Weick, K. E. (1995). What Theory is Not, Theorizing Is. *Administrative Science Quarterly*, *40*(3), 385–390. https://doi.org/10.2307/2393789

Welser, H., Smith, M., Fisher, D., & Gleave, E. (2010). Distilling digital traces: Computational social science approaches to studying the internet. In N. G. Fiedling, R. M. Lee, & G. Blank (Eds.), *The Sage Handbook of Online Research Methods* (pp. 116–140). SAGE.

Weltevrede, E., & Borra, E. (2016). Platform affordances and data practices: The value of dispute on Wikipedia. *Big Data & Society*, *3*(1), 2053951716653418. https://doi.org/10.1177/2053951716653418

Weltevrede, E., & Jansen, F. (2019). Infrastructures of Intimate Data: Mapping the Inbound and Outbound Data Flows of Dating Apps. *Computational Culture*, *7*. http://computationalculture.net/infrastructures-of-intimate-data-mapping-the-inbound-and-outbound-data-flows-of-dating-apps/

West, S. M. (2019). Data Capitalism: Redefining the Logics of Surveillance and Privacy. *Business & Society*, *58*(1), 20–41. https://doi.org/10.1177/0007650317718185

Whitman, M. (2020). "We called that a behavior": The making of institutional data. *Big Data & Society*, *7*(1), 2053951720932200. https://doi.org/10.1177/2053951720932200

Whittaker, M. (2021). The steep cost of capture. *Interactions*, *28*(6), 50–55. https://doi.org/10.1145/3488666

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18. https://doi.org/10.1145/3351095.3372833

Willems, T., & Hafermalz, E. (2021). Distributed seeing: Algorithms and the reconfiguration of the workplace, a case of 'automated' trading. *Information and Organization*, *31*(4), 100376. https://doi.org/10.1016/j.infoandorg.2021.100376

Williamson, B. (2016). Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, *31*(2), 123–141. https://doi.org/10.1080/02680939.2015.1035758

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice* (1st edition). SAGE Publications.

Williamson, B. (2018, March 17). 10 definitions of datafication (in education). *Code Acts in Education*. https://codeactsineducation.wordpress.com/2018/03/17/10-definitions-datafication/

Williamson, B. (2019). Brain Data: Scanning, Scraping and Sculpting the Plastic Learning Brain Through Neurotechnology. *Postdigital Science and Education*, *1*(1), 65–86. https://doi.org/10.1007/s42438-018-0008-5

Witschge, T., & Harbers, F. (2018). 6. Journalism as Practice. In *Journalism* (Vol. 19). De Gruyter Mouton. https://doi.org/10.1515/9781501500084-006

Witteborn, S. (2021). Digital placemaking and the datafication of forced migrants. *Convergence*, *27*(3), 637–648. https://doi.org/10.1177/13548565211003876

Woodward, S. (2015). Object interviews, material imaginings and 'unsettling' methods: Interdisciplinary approaches to understanding materials and material culture: *Qualitative Research*. https://doi.org/10.1177/1468794115589647

Woodward, S. (2019). *Material Methods: Researching and Thinking with Things*. SAGE.

Zakharova, I. (2021). DATA INQUIRY: METHODOLOGICAL CONSIDERATIONS ON DATAFICATION IN SOCIAL RESEARCH. *AoIR Selected Papers of Internet Research*. https://doi.org/10.5210/spir.v2021i0.12082

Zakharova, I., & Jarke, J. (2022). Educational technologies as matters of care. *Learning, Media and Technology*, *47*(1), 95–108. https://doi.org/10.1080/17439884.2021.2018605

Zakharova, I., Jarke, J., & Breiter, A. (2022). Researching #twitterlehrerzimmer during the Covid-19 pandemic: From affinity space to the process understanding of educational data. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *The Ambivalences of Data Power: New perspectives in critical data studies*. Palgrave.

Zegura, E., DiSalvo, C., & Meng, A. (2018). Care and the Practice of Data Science for Social Good. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 1–9. https://doi.org/10.1145/3209811.3209877

Zhao, L., & Ye, W. (2022). Visualization as infrastructure: China's data visualization politics during COVID-19 and their implications for public health emergencies. *Convergence*, 13548565211069872. https://doi.org/10.1177/13548565211069872

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*.

Appendix 1

Table. A list with numeric and full bibliographic references to the research articles sampled for the literature analysis.

Numeric reference	Full bibliographic reference to the research articles sampled for literature analysis
1	Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2). https://doi.org/10.1177/2053951715594634
2	Baack, S. (2018). Practically Engaged: The entanglements between data journalism and civic tech. *Digital Journalism*, 6(6), 673–692. https://doi.org/10.1080/21670811.2017.1375382
3	Barassi, V. (2019). Datafied Citizens in the Age of Coerced Digital Participation. *Sociological Research Online*, 24(3), 414–429. https://doi.org/10.1177/1360780419857734
4	Bayne, S., Connelly, L., Grover, C., Osborne, N., Tobin, R., Beswick, E., & Rouhani, L. (2019). The social value of anonymity on campus: A study of the decline of Yik Yak. *Learning, Media and Technology*, 44(2), 92–107. https://doi.org/10.1080/17439884.2019.1583672
5	Beer, D. (2017). The data analytics industry and the promises of real-time knowing: Perpetuating and deploying a rationality of speed. *Journal of Cultural Economy*, 10(1), 21–33. https://doi.org/10.1080/17530350.2016.1230771
6	Bradbury, A. (2019). Datafied at four: The role of data in the 'schoolification' of early childhood education in England. *Learning, Media and Technology*, 44(1), 7–21. https://doi.org/10.1080/17439884.2018.1511577
7	Candido, H. H. D. (2020). Datafication in schools: Enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*, 29(1–2), 126–157. https://doi.org/10.1080/09620214.2019.1656101
8	Chen, J. Y., & Qiu, J. L. (2019). Digital utility: Datafication, regulation, labor, and DiDi's platformization of urban transport in China. *Chinese Journal of Communication*, 12(3), 274–289. https://doi.org/10.1080/17544750.2019.1614964
9	Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*, 3(2). https://doi.org/10.1177/2053951716679678
10	Dencik, L., Redden, J., Hintz, A., & Warne, H. (2019). The 'golden view': Data-driven governance in the scoring society. *Internet Policy Review*, 8(2). https://doi.org/10.14763/2019.2.1413
11	Duguay, S. (2018). Social media's breaking news: The logic of automation in Facebook Trending Topics and Twitter Moments. *Media International Australia*, 166(1), 20–33. https://doi.org/10.1177/1329878X17737407
12	Engebretsen, M., Kennedy, H., & Weber, W. (2018). Data Visualization in Scandinavian Newsrooms. *Nordicom Review*, 39(2), 3–18. https://doi.org/10.2478/nor-2018-0007
13	Gerhard, U., & Hepp, A. (2018). Appropriating Digital Traces of Self-Quantification: Contextualizing Pragmatic and Enthusiast Self-Trackers. *International Journal of Communication*, 12, 683–700.
14	Halkort, M. (2019). Decolonizing Data Relations: On the Moral Economy of Data Sharing In Palestinian Refugee Camps. *Canadian Journal of Communication*, 44(3). https://doi.org/10.22230/cjc.2019v44n3a3457
15	Hand, M., & Gorea, M. (2018). Digital Traces and Personal Analytics: ITime, Self-Tracking, and the Temporalities of Practice. *International Journal of Communication*, 12, 666-682.
16	Hardy, I., & Lewis, S. (2018). Visibility, invisibility, and visualisation: The danger of school performance data. *Pedagogy, Culture & Society*, 26(2), 233–248. https://doi.org/10.1080/14681366.2017.1380073

17	Hartong, S., & Förschler, A. (2019). Opening the black box of data-based school monitoring: Data infrastructures, flows and practices in state education agencies. *Big Data & Society*, 6(1). Https://doi.org/10.1177/2053951719853311
18	Heeks, R., & Shekhar, S. (2019). Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society*, 22(7), 992–1011. https://doi.org/10.1080/1369118X.2019.1599039
19	Hill, R. L., Kennedy, H., & Gerrard, Y. (2016). Visualizing Junk: Big Data Visualizations and the Need for Feminist Data Studies. *Journal of Communication Inquiry*, 40(4), 331–350. https://doi.org/10.1177/0196859916666041
20	Jackson, S. K., & Kuehn, K. M. (2016). Open Source, Social Activism and 'Necessary Trade-offs' in the Digital Enclosure: A Case Study of Platform Co-operative, Loomio.org. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 14(2). https://doi.org/10.31269/triplec.v14i2.764
21	Janasik-Honkela, N. (2018). Reclaiming Melancholy by Emotion Tracking? Datafication of Emotions in Health Care and at the Workplace. *Open Cultural Studies*, 1(1), 549–558. https://doi.org/10.1515/culture-2017-0052
22	Johanes, P., & Thille, C. (2019). The heart of educational data infrastructures = Conscious humanity and scientific responsibility, not infinite data and limitless experimentation. *British Journal of Educational Technology*, 50(6), 2959–2973. https://doi.org/10.1111/bjet.12862
23	Kannengießer, S. (2020). Reflecting and acting on datafication – CryptoParties as an example of re-active data activism. Convergence: *The International Journal of Research into New Media Technologies*, 26(5–6), 1060–1073. https://doi.org/10.1177/1354856519893357
24	Kelly, S., & Noonan, C. (2017). The Doing of Datafication (And What this Doing Does): Practices of Edification and the Enactment of New Forms of Sociality in the Indian Public Health Service. *Journal of the Association for Information Systems*, 18(12), 872–899. https://doi.org/10.17705/1jais.00477
25	Kennedy, H., & Hill, R. L. (2018). The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation. *Sociology*, 52(4), 830–848. https://doi.org/10.1177/0038038516674675
26	Lee, C. S. (2019). Datafication, dataveillance, and the social credit system as China's new normal. *Online Information Review*, 43(6), 952–970. https://doi.org/10.1108/OIR-08-2018-0231
27	Lehtiniemi, T. (2017). Personal Data Spaces: An Intervention in Surveillance Capitalism? *Surveillance & Society*, 15(5), 626–639. https://doi.org/10.24908/ss.v15i5.6424
28	Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. *Big Data & Society*, 6(1). https://doi.org/10.1177/2053951718821146
29	Leurs, K. (2017). Feminist Data Studies: Using Digital Methods for Ethical, Reflexive and Situated Socio-Cultural Research. *Feminist Review*, 115(1), 130–154. https://doi.org/10.1057/s41305-017-0043-1
30	Maasø, A., & Hagen, A. N. (2020). Metrics and decision-making in music streaming. *Popular Communication*, 18(1), 18–31. https://doi.org/10.1080/15405702.2019.1701675
31	Macgilchrist, F. (2019). Cruel optimism in edtech: When the digital data practices of educational technology providers inadvertently hinder educational equity. *Learning, Media and Technology*, 44(1), 77–86. https://doi.org/10.1080/17439884.2018.1556217
32	Masiero, S., & Das, S. (2019). Datafying anti-poverty programmes: Implications for data justice. *Information, Communication & Society*, 22(7), 916–933. https://doi.org/10.1080/1369118X.2019.1575448
33	Neumann, E. (2019). Setting by numbers: Datafication processes and ability grouping in an English secondary school. *Journal of Education Policy*, 36(1), 1–23. https://doi.org/10.1080/02680939.2019.1646322

34	Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Serrano, J. C. M. (2018). Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*, 5(2). https://doi.org/10.1177/2053951718811844
35	Parisi, L., & Comunello, F. (2020). Dating in the time of "relational filter bubbles": Exploring imaginaries, perceptions and tactics of Italian dating app users. *The Communication Review*, 23(1), 66–89. https://doi.org/10.1080/10714421.2019.1704111
36	Piattoeva, N. (2016). The imperative to protect data and the rise of surveillance cameras in administering national testing in Russia. *European Educational Research Journal*, 15(1), 82–98. https://doi.org/10.1177/1474904115617767
37	Pierlejewski, M. (2020). The data-doppelganger and the cyborg-self: Theorising the datafication of education. *Pedagogy, Culture & Society*, 28(3), 463–475. https://doi.org/10.1080/14681366.2019.1653357
38	Porlezza, C., & Splendore, S. (2019). From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*, 7(9), 1230–1252. https://doi.org/10.1080/21670811.2019.1657778
39	Pybus, J., Coté, M., & Blanke, T. (2015). Hacking the social life of Big Data. *Big Data & Society*, 2(2). https://doi.org/10.1177/2053951715616649
40	Ratner, H., Andersen, B. L., & Madsen, S. R. (2019). Configuring the teacher as data user: Public-private sector mediations of national test data. *Learning, Media and Technology*, 44(1), 22–35. https://doi.org/10.1080/17439884.2018.1556218
41	Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*, 5(2). https://doi.org/10.1177/2053951718809145
42	Roberts-Holmes, G. (2015). The 'datafication' of early years pedagogy: 'If the teaching is good, the data should be good and if there's bad teaching, there is bad data'. *Journal of Education Policy*, 30(3), 302–315. https://doi.org/10.1080/02680939.2014.924561
43	Roberts-Holmes, G., & Bradbury, A. (2016a). The datafication of early years education and its impact upon pedagogy. *Improving Schools*, 19(2), 119–128. https://doi.org/10.1177/1365480216651519
44	Roberts-Holmes, G., & Bradbury, A. (2016b). Governance, accountability and the datafication of early years education in England. *British Educational Research Journal*, 42(4), 600–613. https://doi.org/10.1002/berj.3221
45	Saariketo, M. (2019). Encounters with Self-Monitoring Data on ICT Use. Nordicom Review, 40(s1), 125–140. https://doi.org/10.2478/nor-2019-0018
46	Takayama, K., & Lingard, B. (2019). Datafication of schooling in Japan: An epistemic critique through the 'problem of Japanese education'. *Journal of Education Policy*, 34(4), 449–469. https://doi.org/10.1080/02680939.2018.1518542
47	Taylor, L., & Richter, C. (2017). The Power of Smart Solutions: Knowledge, Citizenship, and the Datafication of Bangalore's Water Supply. *Television & New Media*, 18(8), 721–733. https://doi.org/10.1177/1527476417690028
48	Thompson, G., & Cook, I. (2017). The politics of teaching time in disciplinary and control societies. *British Journal of Sociology of Education*, 38(1), 26–37. https://doi.org/10.1080/01425692.2016.1234365
49	Trevisan, F., Bello, B., Vaughan, M., & Vromen, A. (2020). Mobilizing personal narratives: The rise of digital "story banking" in U.S. grassroots advocacy. *Journal of Information Technology & Politics*, 17(2), 146–160. https://doi.org/10.1080/19331681.2019.1705221
50	Walsh, L. (2019). Whose risk and wellbeing? Three perspectives of online privacy in relation to children and young people's wellbeing. *Media International Australia*, 170(1), 90–99. https://doi.org/10.1177/1329878X19828384
51	Winter, C. (2017). Curriculum policy reform in an era of technical accountability: 'Fixing' curriculum, teachers and students in English schools. *Journal of Curriculum Studies*, 49(1), 55–74. https://doi.org/10.1080/00220272.2016.1205138

Appendix 2

Table. List of academic journals in which sampled research articles have been published and a number of articles in the final sample published in respective journal.

Journal	N of articles in the sample
Big Data & Society	7
British Educational Research Journal	1
British Journal of Educational Technology	1
British Journal of Sociology of Education	1
Canadian Journal of Communication	1
Chinese Journal of Communication	1
Communication Review	1
Convergence	1
Digital Journalism	2
European Educational Research Journal	1
Feminist Review	1
Improving Schools	1
Information Communication & Society	2
International Journal of Communication	2
International Studies in Sociology of Education	1
Internet Policy Review	1
Journal of Communication Inquiry	1
Journal of Cultural Economy	1
Journal of Curriculum Studies	1
Journal of Education Policy	3
Journal of Information Technology and Politics	1
Journal of the Association of Information Systems	1
Learning, Media And Technology	4
Media International Australia	2
Nordicom Review	2
Online Information Review	1
Open Cultural Studies	1
Pedagogy Culture And Society	2
Popular Communication	1
Sociological Research Online	1
Sociology-the Journal of the British Sociological Association	1
Surveillance And Society	1
TripleC	1
Television & New Media	1

Appendix 3

An initial sample of references for the literature analysis before any inclusion and exclusion criteria have been applied. The sample includes results of the search queries in the Scopus and WoS databases, duplicates found in both queries have been removed from this list. The references include bibliographic information available at the date of search query, therefore some of the academic articles published online first may lack data about respective journal's volume and issue.

N	Full reference
1	Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; Blos-Heredero, Carmen De-Pa. 2019. Meloda 5: A metric to assess open data reusability. *Profesional de la Informacion*. 28(6). 10.3145/epi.2019.nov.20
2	Agostinho, Daniela. 2018. Chroma key dreams: Algorithmic visibility, fleshy images and scenes of recognition. *PHILOSOPHY OF PHOTOGRAPHY*. 9(2). 131-155. 10.1386/pop.9.2.131_1
3	Agostinho, Daniela. 2019. Archival encounters: rethinking access and care in digital colonial archives. *Archival Science*. 19(2), SI. 141-165. 10.1007/s10502-019-09312-0
4	Agostinho, Daniela. 2019.The optical unconscious of Big Data: Datafication of vision and care for unknown futures. *Big Data & Society*. 6(1). 10.1177/2053951719826859
5	Airoldi, Massimo. Digital traces of taste: methodological pathways for consumer research. *CONSUMPTION MARKETS & CULTURE*. 10.1080/10253866.2019.1690998
6	Allen, W.L., and B.A. Vollmer. 2018. Clean Skins: Making the e-Border Security Assemblage. *Environment and Planning D: Society and Space*. https://doi.org/10.1177/0263775817722565
7	Altwicker, Tilmann. 2019. International Legal Scholarship and the Challenge of Digitalization. *CHINESE JOURNAL OF INTERNATIONAL LAW*. 18(2). 217-246. 10.1093/chinesejil/jmz012
8	Andrew, D. 2019. Programmatic Trading: The Future of Audience Economics. *Communication Research and Practice.* https://doi.org/10.1080/22041451.2019.1561398
9	Arora, Payal. 2019. Decolonizing Privacy Studies. ***TEL****EVISION & NEW MEDIA*. 20(4), SI. 366-378. 10.1177/1527476418806092
10	Arsenault, Amelia H. 2017. The datafication of media: Big data and the media industries. *INTERNATIONAL JOURNAL OF MEDIA & CULTURAL POLITICS*. 13(1-2). 7-24. 10.1386/macp.13.1-2.7_1
11	Atenas, Javiera; Havemann, Leo; Nascimbeni, Fabio; Villar-Onrubia, Daniel; Orlic, Davor. 2019. Fostering Openness in Education: Considerations for Sustainable Policy-Making. *OPEN PRAXIS*. 11(2). 167-183. 10.5944/openpraxis.11.2.947
12	Aunspach, Chase. .Discrete and looking (to profit): homoconnectivity on Grindr*. *Critical Studies in Media Communication*. 10.1080/15295036.2019.1690157
13	Baack, Stefan. 2015. Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*. 2(2). 10.1177/2053951715594634
14	Baack, Stefan. 2018. PRACTICALLY ENGAGED The entanglements between data journalism and civic tech. *Digital Journalism*. 6(6). 673-692. 10.1080/21670811.2017.1375382

15	Barassi, V. 2016. Datafied Citizens? Social Media Activism, Digital Traces and the Question about Political Profiling. *Communication and the Public*. https://doi.org/10.1177/2057047316683200
16	Barassi, Veronica. 2019. Datafied Citizen in the Ag of Coerced Digital Participation. *Sociological Research Online*. 24(3). 414-429. 10.1177/1360780419857734
17	Barbou, Joshua B. Paperwork. *Health Communication*. 10.1080/10410236.2019.1613481.
18	Bauer, Susanne. 2019. Indexing, Codin}, Scorin}: The Engine Room of Epidemiology and its Routinized Techno-Digestions. *Somatechnics*. 9(2). 223-243. 10.3366/soma.2019.0281
19	Bayne, Sian; Connelly, Louise; Grover, Claire; Osborne, Nicola; Tobin, Richard; Beswick, Emily; Rouhani, Lilinaz. 2019. The social value of anonymity on campus: a study of the decline of Yik Yak. *Learning, Media and Technology*. 44(2). 92-107. 10.1080/17439884.2019.1583672
20	Beer, David. 2017. The data analytics industry and the promises of real-time knowing: perpetuating and deploying a rationality of speed. *Journal of Cultural Economy*. 10(1). 21-33. 10.1080/17530350.2016.1230771
21	Benezra, Amber. 2016. Datafying microbes: Malnutrition at the intersection of genomics and global health. *BioSocieties*. 11(3), SI. 334-351. 10.1057/biosoc.2016.16
22	Beraldo, Davide; Milan, Stefania. 2019. From data politics to the contentious politics of data. *Big Data & Society*. 6(2). 10.1177/2053951719885967
23	Bibri, S. E. 2019. Data-driven smart sustainable urbanism: the intertwined societal factors underlying its materialization, success, expansion, and evolution. *GeoJournal*. 10.1007/s10708-019-10061-x
24	Bocking, Paul. 2019. The Mexican teachers' movement in the context of neoliberal education policy and strategies for resistance. *JOURNAL OF LABOR AND SOCIETY*. 10.1111/wusa.12380. 22(1). 61-76
25	Bossen, Claus; Chen, Yunan; Pine, Kathleen H. 2019.The emergence of new data work occupations in healthcare: The case of medical scribes. *INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS*. 123. 76-83. 10.1016/j.ijmedinf.2019.01.001
26	Bossetta, Michael. 2018. The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 US Election. *JOURNALISM & MASS COMMUNICATION QUARTERLY*. 95(2). 471-496. 10.1177/1077699018763307
27	Bouhlila, D.S., and F. Sellaouti. 2013. Multiple Imputation Using Chained Equations for Missing Data in TIMSS: A Case Study. *Large-Scale Assessments in Education*. https://doi.org/10.1186/2196-0739-1-4
28	Bradbury, Alice. 2019. Datafied at four: the role of data in the `schoolification' of early childhood education in England. *Learning, Media and Technology*. 44(1), SI. 7-21. 10.1080/17439884.2018.1511577
29	Brannon, Monica M. 2013. Standardized Spaces: Satellite Imagery in the Age of Big Data. *Configurations*. 21(3). 271-299. 10.1353/con.2013.0021
30	Brannon, Monica M. .Datafied and Divided: Techno-Dimensions of Inequality in American Cities. *City & Community*. 16(1).
31	Burns, R. 2018. Datafying Disaster: Institutional Framings of Data Production Following Superstorm Sandy. *Annals of the American Association of Geographers*. https://doi.org/10.1080/24694452.2017.1402673

32	Calvo, P. 2019. The ethics of Smart City (EoSC): moral implications of hyperconnectivity, algorithmization and the datafication of urban digital society. *Ethics and Information Technology*. 10.1007/s10676-019-09523-0
33	Candido, Helena Hinke Dobrochinski. .Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*. 10.1080/09620214.2019.1656101
34	Cantens, Thomas. 2018. The Political Arithmetic of Borders: Towards an Enlightened Form of Criticism. *Global Trade and Customs Journal*. 13(3). 116-127
35	Castro, Robin. 2019. Blended learning in higher education: Trends and capabilities. *Education and Information Technologies*. 24(4). 2523-2546. 10.1007/s10639-019-09886-3
36	Chan, N.K., and L. Humphreys. 2018. Mediatization of Social Space and the Case of Uber Drivers. *Media and Communication*. https://doi.org/10.17645/mac.v6i2.1316
37	Chandler, David. 2015. A World without Causation: Big Data and the Coming of Age of Posthumanism. *MILLENNIUM-JOURNAL OF INTERNATIONAL STUDIES*. 43(3). 833-851. 10.1177/0305829815576817
38	Charitsis, Vassilis. 2019. Survival of the (Data) Fit: Self-Surveillance, Corporate Wellness, and the Platformization of Healthcare. *Surveillance & Society*. 17(1-2), SI. 139-144. 10.24908/ss.v17i1/2.12942
39	Charitsis, Vassilis; Zwick, Detlev; Bradshaw, Alan. 2018. Creating Worlds that Create Audiences: Theorising Personal Data Markets in the Age of Communicative Capitalism. *TRIPLEC-COMMUNICATION CAPITALISM & CRITIQUE*. 16(2). 820-834. 10.31269/triplec.v16i2.1041
40	Chen, Julie Yujie; Qiu, Jack Linchuan. 2019. Digital utility: Datafication, regulation, labor, and DiDi's platformization of urban transport in China. *Chinese Journal of Communication*. 12(3), SI. 274-289. 10.1080/17544750.2019.1614964
41	Cheng, Shenghui, Wen Zhong, Katherine E. Isaacs, and Klaus Mueller. 2018. Visualizing the Topology and Data Traffic of Multi-Dimensional Torus Interconnect Networks. IEEE Access. 10.1109/ACCESS.2018.2872344
42	Cinnamon, Jonathan. Attack the Data: Agency, Power, and Technopolitics in South African Data Activism. *ANNALS OF THE AMERICAN ASSOCIATION OF GEOGRAPHERS*. 10.1080/24694452.2019.1644991
43	Couldry, Nick .2018. Tracing Capitalism's Turn to Data: Or, Contextualizing Daily Life}'s New Data "Context" Commentary. *International Journal of Communication*. .12. 701-705
44	Couldry, Nick; Hepp, Andreas. 2018. The continuing lure of the mediated centre in times of deep mediatization: Media Events and its enduring legacy. *Media, Culture & Society*. 40(1). 114-117. 10.1177/0163443717726009
45	Couldry, Nick; Yu, Jun. 2018. Deconstructing datafication's brave new world. *New Media & Society*. 20(1)2. 4473-4491. 10.1177/1461444818775968
46	Das, Ranjana. 2019. A Field in Flux: The Intriguing Past and the Promising Future of Audience Analysis. ***TEL****EVISION & NEW MEDIA*. 20(2), SI. 123-129. 10.1177/1527476418814592
47	Davraz, Metin, Hilmi Cenk Bayrakci, and Yusuf Yusufoglu. 2016. The Effect of Fiber, Opacifier Ratios and Compression Pressure on the Thermal Conductivity of Fumed Silica Based Vacuum Insulation Panels. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-016-2031-8
48	Delfanti, A. 2019. Machinic Dispossession and Augmented Despotism: Digital Work in an Amazon Warehouse. *New Media and Society*. https://doi.org/10.1177/1461444819891613

49	Dencik, Lina; Hintz, Arne; Cable, Jonathan. 2016. Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*. 10.1177/2053951716679678. 3(2). 1-12
50	Dencik, Lina; Redden, Joanna; Hintz, Arne; Warne, Harry. 2019. The `golden view': data-driven governance in the scoring society. *Internet Policy Review*. 10.14763/2019.2.1413. 8(2).
51	Devictor, Vincent; Bensaude-Vincent, Bernadette. 2016. From ecological records to big data: the invention of global biodiversity. *HISTORY AND PHILOSOPHY OF THE LIFE SCIENCES*. 10.1007/s40656-016-0113-2. 38(4).
52	Diaz-Bone, Rainer. 2019. Statistical Panopticism and Its Critique. *HISTORICAL SOCIAL RESEARCH-HISTORISCHE SOZIALFORSCHUNG*. 44(2), SI. 77-102. 10.12759/hsr.44.2019.2.77-102
53	Dourish, Paul; Cruz, Edgar Gomez. 2018. Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*. 5(2). 10.1177/2053951718784083
54	Duggan, M. 2019. Spatial Media and Cycling Spaces: A Theory of Coded Attractors. *Area*. 10.1111/area.12572
55	Duguay, Stefanie. 2018. Social media's breaking news: the logic of automation in Facebook Trending Topics and Twitter Moments. *Media International Australia*. 166(1). 20-33. 10.1177/1329878X17737407
56	Egbert, S.; Krasmann, S. 2019. Predictive policing: not yet, but soon preemptive? *Policing and Society*. 10.1080/10439463.2019.1611821
57	Egbert, Simon. 2019. Predictive Policing and the Platformization of Police Work. *Surveillance & Society*. 17(1-2), SI. 83-88. 10.24908/ss.v17i1/2.12920
58	Engebretsen, Martin; Kennedy, Helen; Weber, Wibke. 2018. Data Visualization in Scandinavian Newsrooms Emerging Trends in Journalistic Visualization Practices. *NORDICOM REVIEW*. 39(2). 3-18. 10.2478/nor-2018-0007
59	Erkan, Ekin. 2019. PSYCHOPOWER AND ORDINARY MADNESS: RETICULATED DIVIDUALS IN COGNITIVE CAPITALISM.COSMOS AND HISTORY. *THE JOURNAL OF NATURAL AND SOCIAL PHILOSOPHY*. 15(1). 214-241
60	Figlio, D., K. Karbownik, and K. Salvanes. 2017. The Promise of Administrative Data in Education Research. *Education Finance and Policy*. https://doi.org/10.1162/EDFP_a_00229
61	Fiske, A., A. Buyx, and B. Prainsack. 2019. Health Information Counselors: A New Profession for the Age of Big Data. *Academic Medicine*. 10.1097/ACM.0000000000002395
62	Fitzgerald, Andrew Arthur. 2019. "Mapping" Media Spaces: Smoothness, Striation, and the Expropriation of Desire in American Journalism from Postindustrial to Datafied Capitalism.. *Communication Theory*. 10.1093/ct/qtz001
63	Flyverbom, M.; Murray, J. 2018. Datastructuring—Organizing and curating digital traces into action. *Big Data and Society*. 5(2). 10.1177/2053951718799114
64	Flyverbom, Mikkel. 2016. Disclosing and concealing: internet governance, information control and the management of visibility. *Internet Policy Review*. 5(3), SI. 10.14763/2016.3.428
65	Forenbacher, I., S. Husnjak, I. Cvitić, and I. Jovović. 2019. Determinants of Mobile Phone Ownership in Nigeria. *Telecommunications Policy*. https://doi.org/10.1016/j.telpol.2019.03.001

66	Freidberg, S.. 2019. "Unable to Determine": Limits to Metrical Governance in Agricultural Supply chains. *Science Technology and Human Values*. 10.1177/0162243919870234
67	Gallagher, R.. 2019. Volatile Memories: Personal Data and Post Human Subjectivity in The Aspern Papers, Analogue: A Hate Story and Tacoma. *Games and Culture*. 10.1177/1555412019841477.
68	Galliers, R. D.; Newell, S.; Shanks, G.; Topi, H.. 2017. Datification and its human, organizational and societal effects: The strategic opportunities and challenges of algorithmic decision-making. *JOURNAL OF STRATEGIC INFORMATION SYSTEMS*. 26(3), SI. 185-190. 10.1016/j.jsis.2017.08.002
69	Gekker, A.; Hind, S. 2019. Infrastructural surveillance. *New Media and Society*. 10.1177/1461444819879426
70	Gerhard, Ulrike; Hepp, Andreas. 2018. Appropriating Digital Traces of Self-Quantification: Contextualizing Pragmatic and Enthusiast Self-Trackers. *International Journal of Communication*. 12. 683-700
71	Gidaris, C. 2019. Surveillance capitalism, datafication, and unwaged labour: The rise of wearable fitness devices and interactive life insurance. *Surveillance and Society*. 17(1-2). 132-138. 10.24908/ss.v17i1/2.12913
72	Gilmore, James N. 2016. Everywear: The quantified self and wearable fitness technologies. *New Media & Society*. 18(11). 2524-2539. 10.1177/1461444815588768
73	Gray, J. 2016. Datafication and democracy recalibrating digital information systems to address broader societal interests. *Juncture*. 23(3). 197-201. 10.1111/newe.12013
74	Gulson, Kalervo N.; Sellar, Sam. 2019. Emerging data infrastructures and the new topologies of education policy. *ENVIRONMENT AND PLANNING D-SOCIETY & SPACE*. 37(2). 350-366. 10.1177/0263775818813144
75	Gurumurthy, A.. 2018. Where Is the 'Struggle' in Communications for Social Progress? *Global Media and Communication*. https://doi.org/10.1177/1742766518776686
76	Gutiérrez, M. 2019. Participation in a datafied environment: Questions about data literacy. *Comunicacao e Sociedade*. 36. 37-55. 10.17231/comsoc.36(2019).2342
77	Halkort, M. 2019.Decolonizing data relations: On the moral economy of data sharing in Palestinian refugee camps. *Canadian Journal of Communication*. 44(3). 317-329. 10.22230/cjc.2019v44n3a3457
78	Hammond, Philip. 2017. From computer-assisted to data-driven: Journalism and Big Data. *Journalism*. 18(4). 408-424. 10.1177/1464884915620205
79	Hand, Martin; Gorea, Michelle. 2018. Digital Traces and Personal Analytics: ïTime, Self-Tracking, and the Temporalities of Practice. *International Journal of Communication*. 12. 666-682
80	Hansen, Hans Krause. 2015. Numerical operations, transparency illusions and the datafication of governance. *European Journal of Social Theory*. 18(2), SI. 203-220. 10.1177/1368431014555260
81	Hardy, Ian; Lewis, Steven. 2018. Visibility, invisibility, and visualisation: the danger of school performance data. *PEDAGOGY CULTURE AND SOCIETY*. 26(2). 233-248. 10.1080/14681366.2017.1380073
82	Hartong, Sigrid; Foerschler, Annina. 2019. Opening the black box of data-based school monitoring: Data infrastructures, flows and practices in state education agencies. *Big Data & Society*. 6(1). 10.1177/2053951719853311

83	Hartong, Sigrid; Piattoeva, Nelli. Contextualizing the datafication of schooling - a comparative discussion of Germany and Russia. *Critical Studies in Education*. 1-16. 10.1080/17508487.2019.1618887
84	Heeks, Richard; Shekhar, Satyarupa. 2019. Datafication, development and marginalised urban communities: an applied data justice framework. *Information, Communication & Society*. 22(7), SI. 992-1011. 10.1080/1369118X.2019.1599039
85	Helles, R., and M. Flyverbom. 2019. Meshes of Surveillance, Prediction, and Infrastructure: On the Cultural and Commercial Consequences of Digital Platforms. Surveillance and Society. https://doi.org/10.24908/ss.v17i1/2.13120
86	Hepp, A.; Breiter, A.; Friemel, T. N. 2018. Digital traces in context: An introduction. *International Journal of Communication*. 12. 439-449
87	Hesmondhalgh, David; Lobato, Ramon. 2019. Television device ecologies, prominence and datafication: the neglected importance of the set-top box. *Media, Culture & Society*. 41(7), SI. 958-974. 10.1177/0163443719857615
88	Hesselberth, Pepita. 2017. Creative Control: digital labour, superimposition, datafication, and the image of uncertainty. *Digital Creativity*. 28(4), SI. 332-347. 10.1080/14626268.2017.1378686
89	Hill, Rosemary Lucy; Kennedy, Helen; Gerrard, Ysabel. 2016. Visualizing Junk: Big Data Visualizations and the Need for Feminist Data Studies. *Journal of Communication Inquiry*. 40(4). 331-350. 10.1177/0196859916666041
90	Hintz, A.; Dencik, L.; Wahl-Jorgensen, K.. 2017. Digital citizenship and surveillance society: Introduction. *International Journal of Communication*. 11. 731-739
91	Hoeyer, Klaus. 2019. Data as promise: Reconfiguring Danish public health through personalized medicine. *Social Studies of Science*. 49(4), SI. 531-555. 10.1177/0306312719858697
92	Hogle, Linda F. 2019.Accounting for accountable care: Value-based population health management. *Social Studies of Science*. 49(4), SI. 556-582. 10.1177/0306312719840429
93	Holloway, Donell. 2019. Surveillance capitalism and children's data: the Internet of toys and things for children. *Media International Australia*. 170(1). 27-36. 10.1177/1329878X19828205
94	Holloway, Donell, and Lelia Green. 2016. The Internet of Toys. *Communication Research and Practice*. 10.1080/22041451.2016.1266124
95	Holtzhausen, Derina. 2016. Datafication: threat or opportunity for communication in the public sphere? *Journal of Communication Management*. 20(1). 21-36. 10.1108/JCOM-12-2014-0082
96	Hosseinipour, Arya, Azuraien Bt Japper-Jaafar, and Suzana Yusup. 2019. Calculation of Wax Appearance Temperature Directly from Hydrocarbon Compositions of Crude Oil. *International Journal of Advanced and Applied Sciences*. https://doi.org/10.21833/ijaas.2019.01.012
97	Hulsey, N.; Reeves, J..2014.The gift that keeps on giving: Google}, ingress, and the gift of surveillance. *Surveillance and Society*. 12(3). 389-400
98	Hursh, David. 2017. Education policy, globalization, commercialization: An interview with Bob Lingard by David Hursh. *Policy Futures in Education*. 15(4), SI. 526-536. 10.1177/1478210317715800
99	Iliadis, Andrew. 2018. Algorithms, ontology, and social progress. *Global Media and Communication*. 14(2). 219-230. 10.1177/1742766518776688

100	Iveson, K., and S. Maalsen. 2019. Social Control in the Networked City: Datafied Dividuals, Disciplined Individuals and Powers of Assembly. *Environment and Planning D: Society and Space*. https://doi.org/10.1177/0263775818812084
101	Jackson, S.; Kuehn, K. M. 2016. Open source, social activism and "Necessary trade-offs" in the digital enclosure: A case study of platform cooperative, Loomio.org. *TripleC*. 14(2). 413-427
102	Jameson, Shazade; Richter, Christine; Taylor, Linnet. 2019. People's strategies for perceived surveillance in Amsterdam Smart City. *Urban Geography*. 40(1)0. 1467-1484. 10.1080/02723638.2019.1614369
103	Janasik-Honkela, Nina. 2017. Reclaiming Melancholy by Emotion Tracking? Datafication of Emotions in Health Care and at the Workplace. *OPEN CULTURAL STUDIES*. 1(1). 549-558. 10.1515/culture-2017-0052
104	Jarke, J.; Breiter, A. 2019.Editorial: the datafication of education. *Learning, Media and Technology*. 44(1). 1-6. 10.1080/17439884.2019.1573833
105	Jensen, Klaus Bruhn. 2018. The Double Hermeneutics of Communication Research. *JAVNOST-THE PUBLIC*. 25(1-2), SI. 177-183. 10.1080/13183222.2018.1418968
106	Jensen, Klaus Bruhn. 2019. The Double Hermeneutics of Audience Research. ***TEL****EVISION & NEW MEDIA*. 20(2), SI. 142-154. 10.1177/1527476418811103
107	Jesse, Norbert. 2018. Internet of Things and Big Data: the disruption of the value chain and the rise of new software ecosystems. *AI & SOCIETY*. 33(2), SI. 229-239. 10.1007/s00146-018-0807-y
108	Johanes, Petr; Thille, Candace. 2019. The heart of educational data infrastructures-Conscious humanity and scientific responsibility, not infinite data and limitless experimentation. *British Journal of Educational Technology*. 50(6). 2959-2973. 10.1111/bjet.12862
109	Jones, Kyle M. L.; McCoy, Chase. 2019. Reconsidering data in learning analytics: opportunities for critical research using a documentation studies framework. *Learning, Media and Technology*. 44(1), SI. 52-63. 10.1080/17439884.2018.1556216
110	Jones, Matthew. 2019. What we talk about when we talk about (big) data. *JOURNAL OF STRATEGIC INFORMATION SYSTEMS*. 28(1). 3-16. 10.1016/j.jsis.2018.10.005
111	Kalidindi, Surya R.; Brough, David B.; Li, Shengyen; Cecen, Ahmet; Blekh, Aleksandr L.; Congo, Faical Yannick P.; Campbell, Carelyn. 2016. Role of materials data science and informatics in accelerated materials innovation. *MRS BULLETIN*. 41(8). 596-602. 10.1557/mrs.2016.164
112	Kannengiesser, Sigrid. Reflecting and acting on datafication - CryptoParties as an example of re-active data activism. *Convergence*. 10.1177/1354856519893357
113	Karlsson, Amanda. 2019. A Room of One's Own? Using Period Trackers to Escape Menstrual Stigma. *Nordicom Review*. https://doi.org/10.2478/nor-2019-0017
114	Katzenbach, C., Baechle, T.C. 2019. Defining concepts of the digital society. *Internet Policy review*. 8(4).
115	Katzenbach, C.; Ulbricht, L. 2019. Algorithmic governance. *Internet Policy Review*. (8).4. 10.14763/2019.4.1424
116	Kear, Mark. 2017. Playing the credit score game: algorithms, "positive' data and the personification of financial objects. *Economy and Society*. 46(3-4). 346-368. 10.1080/03085147.2017.1412642
117	Kelly, Seamas; Noonan, Camilla. 2017. The Doing of Datafication (And What this Doing Does): Practices of Edification and the Enactment of New Forms of Sociality

	in the Indian Public Health Service. *Journal of the Association of Information Systems*. 18(12), SI. 872-899. 10.17705/1jais.00477
118	Kennedy, Helen; Bates, Jo. 2017. Data Power in Material Contexts: Introduction. *TELEVISION & NEW MEDIA*. 18(8), SI. 701-705. 10.1177/1527476417720034
119	Kennedy, Helen; Hill, Rosemary Lucy. 2017. The Pleasure and Pain of Visualizing Data in Times of Data Power}. ***TEL****EVISION & NEW MEDIA*. 18(8), SI. 769-782. 10.1177/1527476416667823
120	Kennedy, Helen; Hill, Rosemary Lucy. 2018. The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation. *SOCIOLOGY-THE JOURNAL OF THE BRITISH SOCIOLOGICAL ASSOCIATION*. 52(4). 830-848. 10.1177/0038038516674675
121	Kennedy, Helen; Poell, Thomas; van Dijck, Jose. 2015. Data and agency. *Big Data & Society*. 2(2). 10.1177/2053951715621569
122	Kitchin, Rob. 2016. The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 374. 2083. 10.1098/rsta.2016.0115
123	Knox, Jeremy; Williamson, Ben; Bayne, Sian. 2020. Machine behaviourism: future visions of `learnification' and `datafication' across humans and digital technologies. *Learning, Media and Technology*. 45(1), SI. 31-45. 10.1080/17439884.2019.1623251
124	Kuntsman, Adi; Miyake, Esperanza; Martin, Sam.2019.Re-thinking Digital Health: Data, Appisation and the (im)possibility of `Opting out'. *DIGITAL HEALTH*. 5. 10.1177/2055207619880671
125	Langlois, Ganaele; Elmer, Greg. 2019. Impersonal subjectivation from platforms to infrastructures. *Media, Culture & Society*. 41(2), SI. 236-251. 10.1177/0163443718818374
126	Larsson, Stefan. 2018. Algorithmic governance and the need for consumer empowerment in data-driven markets. *Internet Policy Review*. 7(2), SI. 10.14763/2018.2.791
127	Leaver, T. 2017. Intimate surveillance: normalizing parental monitoring and mediation of infants online. *Social Media and Society*. 3(2). 10.1177/2056305117707192
128	Leaver, T.; Nansen, B. 2017. Infancy online: An introduction. *Social Media and Society*. 3(2). 10.1177/2056305117707193
129	Lee, C. S. 2019. Datafication, dataveillance, and the social credit system as China's new normal. *Online Information Review*. 43(6). 952-970. 10.1108/OIR-08-2018-0231
130	Lehmann, Joerg; Huber, Elisabeth. 2019. Lost in Datafication? - A Typology of (Emotion) Data Contextualization. *Integrative Psychological and Behavioral Science*. 53(3). 357-373. 10.1007/s12124-018-9470-6
131	Lehmuskallio, Asko; Lampinen, Airi. 2019. Material Mediations Complicate Communication Privacy Management: The Case of Wilma in Finnish High Schools. *International Journal of Communication*. 13. 5752-5770
132	Lehtiniemi, T. 2017. Personal data spaces: An intervention in surveillance capitalism? *Surveillance and Society*. 15(5). 629-639. 10.24908/ss.v15i5.6424
133	Lehtiniemi, T.; Ruckenstein, M. 2019. The social imaginaries of data activism. *Big Data and Society*. 6(1). 10.1177/2053951718821146
134	Lehtiniemi, Tuukka, and Jesse Haapoja. 2020. Data Agency at Stake: MyData Activism and Alternative Frames of Equal Participation. *New Media & Society*. https://doi.org/10.1177/1461444819861955
135	Leicht-Deobald, Ulrich; Busch, Thorsten; Schank, Christoph; Weibel, Antoinette; Schafheitle, Simon; Wildhaber, Isabelle; Kasper, Gabriel. 2019. The Challenges of

	Algorithm-{Based HR Decision-Making for Personal Integrity. *JOURNAL OF BUSINESS ETHICS*. 160(2), SI. 377-392. 10.1007/s10551-019-04204-w
136	Leszczynski, A. 2019. Digital Methods III: The Digital Mundane. *Progress in Human Geography*. https://doi.org/10.1177/0309132519888687
137	Leurs, Koen. 2017. Feminist data studies: using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review*. 115. 130-154. 10.1057/s41305-017-0043-1
138	Lewis, S., and J. Holloway. 2019. Datafying the Teaching 'Profession': Remaking the Professional Teacher in the Image of Data. *Cambridge Journal of Education*. https://doi.org/10.1080/0305764X.2018.1441373
139	Livingstone, Sonia. 2019. Audiences in an Age of Datafication: Critical Questions for Media Research. ***TEL****EVISION & NEW MEDIA*. 20(2), SI. 170-183. 10.1177/1527476418811118
140	Lundie, D. 2016. Authority, Autonomy and Automation: The Irreducibility of Pedagogy to Information Transactions. *Studies in Philosophy and Education*. https://doi.org/10.1007/s11217-016-9517-4
141	Lupton, D. 2018. How Do Data Come to Matter? Living and Becoming with Personal Data. *Big Data and Society*. https://doi.org/10.1177/2053951718786314
142	Lupton, Deborah; Williamson, Ben. 2017. The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*. 19(5), SI. 780-794. 10.1177/1461444816686328
143	Maalsen, Sophia; Sadowski, Jathan. 2019. The Smart Home on FIRE: Amplifying and Accelerating Domestic Surveillance. *Surveillance & Society*. 17(1-2), SI. 118-124. 10.24908/ss.v17i1/2.12925
144	Maasø, A.; Hagen, A. N..2019. Metrics and decisions-making in music streaming. *Popular Communication*. 10.1080/15405702.2019.1701675
145	Macgilchrist, Felicitas. 2019. Cruel optimism in edtech: when the digital data practices of educational technology providers inadvertently hinder educational equity. *Learning, Media and Technology*. 44(1), SI. 77-86. 10.1080/17439884.2018.1556217
146	Madsen, A. K.; Flyverbom, M.; Hilbert, M.; Ruppert, E. 2016. Big data: Issues for an international political sociology of data practices. *International Political Sociology*. 10(3). 275-296. 10.1093/ips/olw010
147	Mai, Jens-Erik. 2016. Big data privacy: The datafication of personal information. *Information Society*. 32(3). 192-199. 10.1080/01972243.2016.1153010
148	Manolev, Jamie; Sullivan, Anna; Slee, Roger. 2019. The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Learning, Media and Technology*. 44(1), SI. 36-51. 10.1080/17439884.2018.1558237
149	Markham, Annette N. 2019.Critical Pedagogy as a Response to Datafication. *Qualitative Inquiry*. 25(8), SI. 754-760. 10.1177/1077800418809470
150	Masiero, Silvia; Das, Soumyo. 2019. Datafying anti-poverty programmes: implications for data justice. *Information, Communication & Society*. 22(7), SI. 916-933. 10.1080/1369118X.2019.1575448
151	McEvoy, Fiona J. 2017.DECISIONS, DECISIONS: BIG DATA AND THE FUTURE OF AUTONOMY. *ANALELE UNIVERSITATII BUCURESTI. FILOSOFIE-ANNALS OF THE UNIVERSITY OF BUCHAREST-PHILOSOPHY*. 66(2). 43-66
152	McFall, Liz; Moor, Liz. 2018. Who, or what, is insurtech personalizing?: persons, prices and the historical classifications of risk. *DISTINKTION-JOURNAL OF SOCIAL THEORY*. 19(2), SI. 193-213. 10.1080/1600910X.2018.1503609

153	McKinson, K. D. 2019. Black carcerality and emancipation in Postcolonial Jamaica. *Surveillance and Society*. 17(5). 734-737. 10.24908/ss.v17i5.13437
154	Mejias, U. A.; Couldry, N. 2019.Datafication. *Internet Policy Review*. 8(4). 10.14763/2019.4.1428
155	Mellet, K.; Beauvisage, T. 2019. Cookie monsters. Anatomy of a digital market infrastructure. *Consumption Markets and Culture*. 10.1080/10253866.2019.1661246
156	Milan, S. 2018. Political agency, digital traces, and bottom-up data practices. *International Journal of Communication*. 12. 507-527
157	Milan, S.; Treré, E. 2019. Big Data from the South(s): Beyond Data Universalism. *Television and New Media*. 20(4). 319-335. 10.1177/1527476419837739
158	Miles, Christopher. 2019. The combine will tell the truth: On precision agriculture and algorithmic rationality. *Big Data & Society*. 6(1). 10.1177/2053951719849444
159	Millei, Zsuzsa; Gallagher, Jannelle. 2017. Ad-hoc numbers forming provision and policy: round and round of universal access in an Australian preschool. *EARLY CHILD DEVELOPMENT AND CARE*. 187(10), SI. 1528-1542. 10.1080/03004430.2017.1289926
160	Millington, B.; Millington, R..2015.'The datafication of everything': Toward a sociology of sport and big data. *Sociology of Sport Journal*. 32(2). 140-160. 10.1123/ssj.2014-0069
161	Minca, C.; Roelofsen, M. 2019. Becoming Airbnbeings: on datafication and the quantified Self in tourism. *Tourism Geographies*. 10.1080/14616688.2019.1686767
162	Murray, S. 2019. Secret agents: Algorithmic culture, Goodreads and datafication of the contemporary book world. *European Journal of Cultural Studies*. 10.1177/1367549419886026
163	Murrell, M. 2014. The datafied book and the entanglements of digitization. *Anthropology Today*. 30(5). 3-6. 10.1111/1467-8322.12129
164	Murschetz, Paul Clemens; Schluetz, Daniela.2018.BIG DATA AND ` TELEVISION BROADCASTING. A CRITICAL REFLECTION ON BIG DATA'S SURGE. TO BECOME A NEW TECHNO-ECONOMIC PARADIGM AND ITS IMPACTS ON THE CONCEPT OF THE "ADDRESSABLE AUDIENCE". *FONSECA-JOURNAL OF COMMUNICATION*. 17. 23-38. 10.14201/fjc2018172338
165	Neumann, Eszter. Setting by numbers: datafication processes and ability grouping in an English secondary school. *Journal of Education Policy*. 10.1080/02680939.2019.1646322
166	Nieborg, D.B., and T. Poell. 2018. The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity. *New Media and Society*. https://doi.org/10.1177/1461444818769694
167	Oertzen, Christine von. 2018. Datafication and Spatial Visualization in Nineteenth-Century Census Statistics. *HISTORICAL STUDIES IN THE NATURAL SCIENCES*. 48(5), SI. 568-580. 10.1525/hsns.2018.48.5.568
168	Orlova, G.; Kasatkina, A. 2017. Wide open qualitative data: the Obninsk digital project as an ethical dispositive. *Russian Journal of Communication*. 9(3). 328-335. 10.1080/19409419.2017.1376846
169	Pangrazio, L.; Sefton-Green, J. 2019. The social utility of 'data literacy'. *Learning, Media and Technology*. 10.1080/17439884.2020.1707223
170	Papakyriakopoulos, Orestis; Hegelich, Simon; Shahrezaye, Morteza; Serrano, Juan Carlos Medina. 2018. Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*. 5(2). 10.1177/2053951718811844

171	Parisi, L.; Comunello, F. 2019. Dating in the time of "relational filter bubbles": exploring imaginaries, perceptions and tactics of Italian dating app users. *Communication Review*. 10.1080/10714421.2019.1704111
172	Peckham, Robert; Sinha, Ria. 2019. Anarchitectures of health: Futures for the biomedical drone. *GLOBAL PUBLIC HEALTH*. 14(8). 1204-1219. 10.1080/17441692.2018.1546335
173	Pellegrino, Giuseppina; Soderberg, Johan; Milan, Stefania. 2019. Datafication from Below: Epistemology, Ambivalences, Challenges. *TECNOSCIENZA-ITALIAN JOURNAL OF SCIENCE & TECHNOLOGY STUDIES*. 10(1). 89-113
174	Peña Gangadharan, S., and J. Niklas. 2019. Decentering Technology in Discourse on Discrimination. *Information Communication and Society*. https://doi.org/10.1080/1369118X.2019.1593484
175	Piattoeva, Nelli. 2016. The imperative to protect data and the rise of surveillance cameras in administering national testing in Russia. *European Educational Research Journal*. 15(1), SI. 82-98. 10.1177/1474904115617767
176	Pierlejewski, M. 2019. The data-doppelganger and the cyborg-self: theorising the datafication of education. Pedagogy, Culture and Society. 10.1080/14681366.2019.1653357
177	Pink, Sarah; Lanzeni, Debora. 2018. Future Anthropology Ethics and Datafication: Temporality and Responsibility in Research. *Social Media & Society*. 4(2). 10.1177/2056305118768298
178	Poell, T.; Nieborg, D.; van Dijck, J. 2019. Platformisation. *Internet Policy Review*. 8(4). 10.14763/2019.4.1425
179	Pold, Soren Bro. 2019. New ways of hiding: towards metainterface realism. *Artnodes*. 10.7238/a.v0i24.3283. 24. 72-82.
180	Ponzanesi, Sandra. 2019. Migration and Mobility in a Digital Age: (Re)Mapping Connectivity and Belonging. ***TEL****EVISION & NEW MEDIA*. 20(6), SI. 547-557. 10.1177/1527476419857687
181	Porlezza, C.; Splendore, S. 2019. From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*. 7(9). 1230-1252. 10.1080/21670811.2019.1657778
182	Prince, R. 2019. The geography of statistics: Social statistics from moral science to big data. *Progress in Human Geography*. 10.1177/0309132519873421
183	Purtova, N. 2018. The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*. 10(1). 40-81. 10.1080/17579961.2018.1452176
184	Pybus, Jennifer; Cote, Mark; Blanke, Tobias. 2015. Hacking the social life of Big Data. *Big Data & Society*. 2(2). 10.1177/2053951715616649
185	Rasekh, Mansour, Hadi Bagheri, Mohammad Hossien Kianmehr, and Shahriar Kouravand. 2017. A Study of Some Physical Properties of Lentil and Wild Oat Weed and Parameters Affecting the Separation of Wild Oat Weed from Lentil by a Gravity Table Separator. *Journal of Agricultural Sciences-Tarim Bilimleri Dergisi*.
186	Ratner, Helene; Anderse, Bjarke Lindso; Madsen, Simon Ryberg. 2019. Configuring the teacher as data user: public-private sector mediations of national test data. *Learning, Media and Technology*. 44(1), SI. 22-35. 10.1080/17439884.2018.1556218
187	Redden, Joanna. 2018. Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*. 5(2). 10.1177/2053951718809145

188	Rhoen, Michiel. 2016. Beyond consent: improving data protection through consumer protection law. Internet Policy Review. 5(1), SI. 10.14763/2016.1.404
189	Rhoen, Michiel. 2017. Rear view mirror, crystal ball: Predictions for the future of data protection law based on the history of environmental protection law. *COMPUTER LAW & SECURITY REVIEW*. 33(5). 603-617. 10.1016/j.clsr.2017.05.010
190	Ricaurte, Paola. 2019. Data Epistemologies, The Coloniality of Power, and Resistance. ***TEL****EVISION & NEW MEDIA*. 20(4), SI. 350-365. 10.1177/1527476419831640
191	Rieder, B. 2017. Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality. *Information Communication and Society*. 10.1080/1369118X.2016.1181195
192	Rinik, C. 2019. Data trusts: more data than trust? The perspective of the data subject in the face of a growing problem. *International Review of Law, Computers and Technology*. 10.1080/13600869.2019.1594621
193	Roberts-Holmes, Guy. 2015. The `datafication' of early years pedagogy: `if the teaching is good, the data should be good and if there's bad teaching, there is bad data'. *Journal of Education Policy*. 30(3), SI. 302-315. 10.1080/02680939.2014.924561
194	Roberts-Holmes, Guy; Bradbury, Alice. 2016. Governance, accountability and the datafication of early years education in England. *British Educational Research Journal*. 42(4). 600-613. 10.1002/berj.3221
195	Roberts-Holmes, Guy; Bradbury, Alice. 2016. The datafication of early years education and its impact upon pedagogy. *Improving Schools*. 19(2), SI. 119-128. 10.1177/1365480216651519
196	Robinson, Sandra. 2018. Databases and Doppelgangers: New Articulations of Power. *Configurations*. 26(4). 411-440. 10.1353/con.2018.0035
197	Saariketo, Minna. 2019. Encounters with Self-Monitoring Data on ICT Use. *NORDICOM REVIEW*. 40(1). 125-140. 10.2478/nor-2019-0018
198	Sadinle, Mauricio. 2017. Bayesian Estimation of Bipartite Matchings for Record Linkage. *Journal of the American Statistical Association*.
199	Sadinle, Mauricio, and Stephen E. Fienberg. 2013. A Generalized Fellegi-Sunter Framework for Multiple Record Linkage With Application to Homicide Record Systems. *Journal of the American Statistical Association*.
200	Sadowski, Jathan. 2019. When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*. 6(1). 10.1177/2053951718820549
201	Sandvik, Kristin Bergtora; Jacobsen, Katja Lindskov; McDonald, Sean Martin. 2017. Do no harm: A taxonomy of the challenges of humanitarian experimentation. *INTERNATIONAL REVIEW OF THE RED CROSS*. 904. 319-344. 10.1017/S181638311700042X.99
202	Schmietow, Bettina; Marckmann, Georg. 2019. Mobile health ethics and the expanding role of autonomy. *MEDICINE HEALTH CARE AND PHILOSOPHY*. 22(4). 623-630. 10.1007/s11019-019-09900-y
203	Schüll, N.D. 2019. The data-based self: self quantification and the data-driven (good) life. *Social research: An International Quarterly*. 86(4).
204	Sellar, Sam; Thompson, Greg. 2016. The Becoming-Statistic: Information Ontologies and Computerized Adaptive Testing in Education. *Cultural Studies ↔ Critical Methodologies*. 16(5), SI. 491-501. 10.1177/1532708616655770
205	Shapiro, Aaron. 2018. Street-level: Google Street View's abstraction by datafication. *New Media & Society*. 20(3). 1201-1219. 10.1177/1461444816687293

206	Sheller, T. 2019. Datafied Urban Governance: Real-Time Data Sustainability, Smart Digital Technologies, and the Citizen-Driven Internet of Things. *Geopolitics, History, and International Relations*. 10.22381/GHIR11220192
207	Sleigh, Joanna. 2018. Experiences of Donating Personal Data to Mental Health Research: An Explorative Anthropological Study. *BIOMEDICAL INFORMATICS INSIGHTS*. 10. 10.1177/1178222618785131
208	Smith, G.J.D., and B. Vonthethoff. 2017. Health by Numbers? Exploring the Practice and Experience of Datafied Health. *Health Sociology Review*. https://doi.org/10.1080/14461242.2016.1196600
209	Smith, Gavin J. D. 2016. Surveillance, Data and Embodiment: On the Work of Being Watched. *BODY & SOCIETY*. 22(2), SI. 108-139. 10.1177/1357034X15623622
210	Soledad Segura, Maria; Waisbord, Silvio. 2019. Between Data Capitalism and Data Citizenship. ***TEL****EVISION & NEW MEDIA*. 20(4), SI. 412-419. 10.1177/1527476419834519
211	Steensen, Steen. 2019. Journalism's epistemic crisis and its solution: Disinformation, datafication and source criticism. *Journalism*. 20(1), SI. 185-189. 10.1177/1464884918809271
212	Stevenson, H. 2017. The "Datafication" of Teaching: Can Teachers Speak Back to the Numbers? *Peabody Journal of Education*. 92(4). 537-557. 10.1080/0161956X.2017.1349492
213	Stiegler, Christian. 2017. The Politics of Immersive Storytelling: Virtual Reality and the Logics of Digital Ecosystems. *INTERNATIONAL JOURNAL OF E-POLITICS*. 8(3), SI. 1-15. 10.4018/IJEP.2017070101
214	Stoilova, M.; Nandagiri, R.; Livingstone, S. 2019. Children's understanding of personal data and privacy online–a systematic evidence mapping. *Information Communication and Society*. 10.1080/1369118X.2019.1657164
215	Strauss, Stefan. 2015. Datafication and the Seductive Power of Uncertainty - Critical Exploration of Big Data Enthusiasm. *INFORMATION*. 6(4). 836-847. 10.3390/info6040836
216	Sumartojo, Shanti; Pink, Sarah; Lupton, Deborah; LaBond, Christine Heyes. 2016. The affective intensities of datafied space. *EMOTION SPACE AND SOCIETY*. 21. 33-40. 10.1016/j.emospa.2016.10.004
217	Swist, Teresa; Collin, Philippa; Third, Amanda. 2019. Children's data journeys beyond the `supply chain': co-curating flows and frictions. *Media International Australia*. 170(1). 68-77. 10.1177/1329878X19828390
218	Szulc, Lukasz. 2019. Profiles, Identities, Data: Making Abundant and Anchored Selves in a Platform Society. Communication Theory. 29(3). 257-276. 10.1093/ct/qty031
219	Taheri, Javid, Young Choon Lee, Albert Y. Zomaya, and Howard Jay Siegel. 2013. A Bee Colony Based Optimization Approach for Simultaneous Job Scheduling and Data Replication in Grid Environments. *Computers & Operations Research*.
220	Takayamaa, Keita; Lingard, Bob. 2019. Datafication of schooling in Japan: an epistemic critique through the `problem of Japanese education'. *Journal of Education Policy*. 34(4). 449-469. 10.1080/02680939.2018.1518542
221	Taylor, L..2017. What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally. *Big Data and Society*. 4(4). 10.1177/2053951717736335
222	Taylor, Linnet; Broeders, Dennis. 2015. In the name of Development: Power, profit and the datafication of the global South. *Geoforum*. 64. 229-237. 10.1016/j.geoforum.2015.07.002

223	Taylor, Linnet; Richter, Christine. 2017. The Power of Smart Solutions: Knowledge, Citizenship, and the Datafication of Bangalore's Water Supply. ***TEL****EVISION & NEW MEDIA*. 18(8), SI. 721-733. 10.1177/1527476417690028
224	Thompson, Greg .2017. Computer adaptive testing, big data and algorithmic approaches to education. *British Journal of Sociology of Education*. 38(6). 827-840. 10.1080/01425692.2016.1158640
225	Thompson, Greg; Cook, Ian. 2017. The logic of data-sense: thinking through Learning Personalisation. *Discourse: Studies in the Cultural Politics of Education*. 38(5). 740-754. 10.1080/01596306.2016.1148833
226	Thompson, Greg; Cook, Ian. 2017. The politics of teaching time in disciplinary and control societies. *British Journal of Sociology of Education*. 38(1), SI. 26-37. 10.1080/01425692.2016.1234365
227	Thompson, Greg; Sellar, Sam. 2018. Datafication, testing events and the outside of thought. *Learning, Media and Technology*. 43(2). 139-151. 10.1080/17439884.2018.1444637
228	Thoutenhoofd, Ernst D. 2018.The Datafication of Learning: Data Technologies as Reflection Issue in the System of Education. Studies in Philosophy and Education. 37(5). 433-449. 10.1007/s11217-017-9584-1
229	Thylstrup, Nanna Bonde. 2019.Data out of place: Toxic traces and the politics of recycling. *Big Data & Society*. 6(2). 10.1177/2053951719875479
230	Till, Christopher. 2019. Creating `automatic subjects': Corporate wellness and self-tracking. *HEALTH*. 23(4), SI. 418-435. 10.1177/1363459319829957
231	Trevisan, F.; Bello, B.; Vaughan, M.; Vromen, A..2019. Mobilizing personal narratives: The rise of digital "story banking" in U.S. grassroots advocacy. *Journal of Information Technology and Politics*. 10.1080/19331681.2019.1705221
232	Tsai, Y.-S., C. Perrotta, and D. Gašević. 2019. Empowering Learners with Personalised Learning Approaches? Agency, Equity and Transparency in the Context of Learning Analytics. *Assessment and Evaluation in Higher Education*. 10.1080/02602938.2019.1676396
233	Turvey, Keith. 2019. Humanising as innovation in a cold climate of [so-called-evidence-based] teacher education. *Journal of Education for Teaching*. 45(1), SI. 15-30. 10.1080/02607476.2019.1550603
234	van de Ven, Inge. 2018. The Monumental Knausgard: Big Data, Quantified Self, and Proust for the Facebook Generation. *NARRATIVE*. 26(3). 320-338. 10.1353/nar.2018.0016
235	van de Wiel, Lucy. 2019. The datafication of reproduction: time-lapse embryo imaging and the commercialisation of IVF. *SOCIOLOGY OF HEALTH & ILLNESS*. 41(1), SI. 193-209. 10.1111/1467-9566.12881
236	van Dijck, J. 2014. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*. 12(2). 197-208
237	van Dijck, J.; Poell, T. 2013. Understanding social media logic. *Media and Communication*. 1(1). 2-14. 10.12924/mac2013.01010002
238	van Dijck, Jose; Poell, Thomas. 2016. Understanding the promises and premises of online health platforms. *Big Data & Society*. 3(1). 10.1177/2053951716654173
239	Vannini, S.; Gomez, R.; Newell, B. C. 2019. "Mind the Five": Guidelines for Data Privacy and Security in Humanitarian Work With Undocumented Migrants and Other Vulnerable Populations. *Journal of the Association for Information Science and Technology*. 10.1002/asi.24317

240	Veel, Kristin. 2018. Make data sing: The automation of storytelling. *Big Data & Society*. 5(1). 10.1177/2053951718756686
241	Wallenburg, Iris; Bal, Roland. 2019. The gaming healthcare practitioner: How practices of datafication and gamification reconfigure care. *HEALTH INFORMATICS JOURNAL*. 25(3), SI. 549-557. 10.1177/1460458218796608
242	Walsh, Lucas. 2019. Whose risk and wellbeing? Three perspectives of online privacy in relation to children and young people's wellbeing. *Media International Australia*. 170(1). 90-99. 10.1177/1329878X19828384
243	Warren, C.D., K. Fowler, D. Speed, and A. Walsh. 2018. The Influence of Social Support on Psychological Distress in Canadian Adults with Bipolar Disorder. *Social Psychiatry and Psychiatric Epidemiology*.
244	Webb, Mary E.; Prasse, Doreen; Phillips, Mike; Kadijevich, Djordje M.; Angeli, Charoula; Strijker, Allard; Carvalho, Ana Amelia; Andresen, Bent B.; Dobozy, Eva; Laugesen, Hans. 2018. Challenges for IT-Enabled Formative Assessment of Complex 21st Century Skills. *TECHNOLOGY KNOWLEDGE AND LEARNING*. 23(3), SI. 441-456. 10.1007/s10758-018-9379-7
245	Webb, P. T.; Sellar, S.; Gulson, K. N. 2019. Anticipating education: governing habits, memories and policy-futures. *Learning, Media and Technology*. 10.1080/17439884.2020.1686015
246	Wielki, J. 2015. THE SOCIAL AND ETHICAL CHALLENGES CONNECTED WITH THE BIG DATA PHENOMENON. *POLISH JOURNAL OF MANAGEMENT STUDIES*. 11(2). 192-202
247	Wiley, L., T.E. Chrzastowski, and S. Baker. 2011. A Domestic Monograph Collection Assessment in Illinois Academic Libraries: What Are We Buying and How Is It Used? *Interlending & Document Supply*.
248	Wilson, D. 2019. Platform Policing and the Real-Time Cop. *Surveillance and Society*. https://doi.org/10.24908/ss.v17i1/2.12958
249	Winter, Christine. 2017. Curriculum policy reform in an era of technical accountability: "fixing' curriculum, teachers and students in English schools. *Journal of Curriculum Studies*. 49(1). 55-74. 10.1080/00220272.2016.1205138
250	Winter, Jenifer Sunrise; Davidson, Elizabeth. 2019. Big data governance of personal health information and challenges to contextual integrity. *Information Society*. 35(1). 36-51. 10.1080/01972243.2018.1542648
251	Wolf, Ralf de; Joye, Stijn. 2019. Control Responsibility: The Discursive Construction of Privacy, Teens, and Facebook in Flemish Newspapers. *International Journal of Communication*. 13. 5505-5524
252	Wood, Mark A. 2019. Algorithmic tyranny: Psycho-Pass, science fiction and the criminological imagination. *CRIME MEDIA CULTURE*. 15(2). 323-339. 10.1177/1741659018774609
253	Ylijoki, Ossi; Porras, Jari. 2016. CONCEPTUALIZING BIG DATA: ANALYSIS OF CASE STUDIES. *INTELLIGENT SYSTEMS IN ACCOUNTING FINANCE & MANAGEMENT*. 23(4). 295-310. 10.1002/isaf.1393
254	Ytre-Arne, Brita; Das, Ranjana. 2019. An Agenda in the Interest of Audiences: Facing the Challenges of Intrusive Media Technologies. *TELEVISION & NEW MEDIA*. 20(2), SI. 184-198. 10.1177/1527476418759604
255	Zulhuda, S.; Azmi, I.M.A.G.; Hakiem, N. 2015. Big data, cloud and bring your own device: How the data protection law addresses the impact of "datafication". *Advanced Science Letters*. 21(10). 3346-3350. 10.1166/asl.2015.6493

EIDESSTATTLICHE ERKLÄRUNG

Hiermit versichere ich, dass ich

1. die eingereichte Dissertation selbständig und ohne unerlaubte fremde Hilfe verfasst,
2. keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und
3. die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Bremen, 31. März 2022 Irina Zakharova