DISSERTATIONES
DE MEDIIS ET
COMMUNICATIONIBUS
UNIVERSITATIS
TARTUENSIS
42

MARIS MÄNNISTE

Big data imaginaries of data pioneers: changed data relations and challenges to agency

DISSERTATIONES DE MEDIIS ET COMMUNICATIONIBUS
UNIVERSITATIS TARTUENSIS

42

DISSERTATIONES DE MEDIIS ET COMMUNICATIONIBUS
UNIVERSITATIS TARTUENSIS
42

MARIS MÄNNISTE

Big data imaginaries of data pioneers: changed data relations and challenges to agency

UNIVERSITY OF TARTU
Press
1632

Institute of Social Studies, University of Tartu

Dissertation accepted in fulfilment of the requirements for the degree of Doctor of Philosophy (in Media and Communication) on March 1, 2022, by the Council of the Institute of Social Studies, University of Tartu.

Supervisors: Associate Professor Anu Masso
Institute of Social Studies, University of Tartu

Ragnar Nurkse Department of Innovation and Governance,
Tallinn University of Technology

Professor Andra Siibak
Institute of Social Studies, University of Tartu

Opponent: Associate Professor Mirko Tobias Schäfer
Utrecht University, The Netherlands

Commencement: April 12, 2022, at the White Hall of the University of Tartu Museum

The publication of this dissertation is granted by the Institute of Social Studies, University of Tartu and by the Doctoral School of Behavioural, Social and Health Sciences, created under the auspices of the European Social Fund. This research was also supported by the Estonian government programme Dora Plus, which was implemented with the support of the EU Regional Development Fund, by the University of Tartu ASTRA PER ASPERA Project (financed by the European Regional Development Fund), and IUT 20-38 (Estonian Ministry of Education and Research).

ISSN 1406-2313
ISBN 978-9949-03-847-3 (print)
ISBN 978-9949-03-848-0 (pdf)

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This dissertation is based on the following four original publications listed in chronological order and referred to by Roman numerals:

I. Männiste, M., Masso, A. (2018). The Role of Institutional Trust in Estonians' Privacy Concerns. Studies of Transition States and Societies, 10 (2), 22−39. http://publications.tlu.ee/index.php/stss/article/view/676/535

II. Masso, A., Männiste, M. & Siibak, A. (2020). 'End of Theory' in the Era of Big Data: Methodological Practices and Challenges in Social Media Studies. Acta Baltica Historiae et Philosophiae Scientiarum, 8 (1), 33−61. 10.11590/abhps.2020.1.02, https://www.ies.ee/bahps/acta-baltica/abhps-8-1/02_Masso-2020-1-02.pdf

III. Männiste, M., Masso, A. (2020). "Three Drops of Blood for the Devil": Data Pioneers as Intermediaries of Algorithmic Governance Ideals. Mediální studia / Media Studies, 14 (1), 55−74. https://www.medialnistudia.fsv.cuni.cz/front.file/download?file=medialni_studia_1_2020%2004%20manniste_masso.pdf

IV. Masso, A., Männiste, M. The datafication of mobility governance: The discourses on tracing mobility through data. *Unpublished manuscript.*

AUTHOR'S CONTRIBUTIONS

STUDY I "The Role of Institutional Trust in Estonians' Privacy Concerns": the author was partially involved in writing the data analysis and predominantly responsible for the theoretical and discussion parts of the manuscript.

STUDY II "'End of Theory' in the Era of Big Data: Methodological Practices and Challenges in Social Media Studies": all of the co-authors contributed to the development of the coding guide. The author was predominantly responsible for coding the articles, the data analysis and was partially involved in writing the other parts of the manuscript.

STUDY III "Three Drops of Blood for the Devil Data Pioneers as Intermediaries of Algorithmic Governance Ideals": the author was partially responsible for collecting the data on Estonian data experts. The author was predominantly responsible for writing the data analysis and discussion, and partially involved in writing the theory and methods parts of the manuscript.

STUDY IV "The datafication of mobility governance: The discourses on tracing mobility through data": The author was partially responsible for collecting the data on Estonian data experts. The author was involved in the qualitative data analysis part of the manuscript, and partially involved in writing other parts of the manuscript.

ACKNOWLEDGEMENTS

When I started my PhD journey, I imagined many different possibilities for this thesis and for the PhD process. Although the end result is much different from what I imagined in the beginning and the process itself has been at times challenging and full of obstacles, it has also been a process of finding the researcher in me, developing a deep love for certain topics and having the opportunity to meet and learn from so many people.

My deepest thanks go to my supervisors, Professor **Andra Siibak** and Associate Professor **Anu Masso**, both of whom have provided me with their continuous support, critical feedback and valuable insights into a fast-moving research area. Without them, this research would not have been possible. I am grateful for both of you for taking me on and offering me an amazing opportunity to learn from you throughout this process. My journey towards my PhD has not been straight or smooth, and especially because of that I really value all the times you have been there to listen, offer your advice and support, and help me through the doubts and fears I have had about this process. I am also deeply grateful that you introduced me to so many opportunities, opening several doors and introducing me to people some of whom have now also grown to be important friends and colleagues. You will always be an inspiration to me and I hope to have many more opportunities to work with both of you.

I have also benefited from the support of many of **my colleagues and peers in the Institute of Social Studies.** I am very thankful for all of the support I have had throughout this process. I value all of the discussions I have had about my research, as well as general support, such as sometimes covering my lectures when I really needed to take time away for writing camps. My thanks also go to the **PhD committee** for their constructive criticism and valuable feedback, which has been helpful in bringing my thesis to its present state.

During my studies, I have also been fortunate to have many opportunities to introduce my work at several conferences and workshops, which has been helpful in gaining a better understanding of this topic. I am especially grateful to Professor **Cornelius Puschmann**, whose research group in the Hans-Bredow Institute (Algorithmed Public Spheres) was so kind as to host me and share their valuable insights on this topic. The time spent working with this research group was really insightful and helped me with my PhD thesis and the preparation of articles.

My journey would be in many ways different without the several workshops and winter and summer schools that I have attended during my studies. I am especially grateful to the organisers of the European Media and Communication (ECREA) doctoral summer school, especially to **Andra Siibak, Pille Pruulmann-Vengerfeldt and Risto Kunelius**, who made this experience one of the best during my studies, offered really valuable advice during the discussions in the doctoral school and offered a valuable publishing opportunity.

During this journey, I have also been thankful for having found a supportive community amongst my fellow PhD students and academic colleagues, and for the friendships that have formed. I am happy that I started my academic journey and my teaching with my dearest friends **Regina Lapp and Katrin Kannukene**. I treasure all the moments, discussions and cake we have shared during these times.

My journey would have been so much different without **Mari-Liis Tikerperi**, with whom I started my PhD studies. I am glad that I have found a friend who supports me not just in my research endeavours, but also by contributing to my immense cat cups collection.

I have also been lucky to share my journey with **Tayfun Kasapoglu** who helped me through the pandemic isolation and has always been there by my side, cheering me on. Thank you for always reminding me to be kind to myself and for sharing this journey with me. My wins, big and small, have always been shared also with **Marit Sukk**. I am amazed by your Q-methodology knowledge and I treasure the friendship we have developed throughout this process. Thank you both for the tight group we have formed, which offers a safe place to share all of the different parts of this current journey, as well as future journeys.

I am also grateful to **Stephen Cory Robinson**, whom I was happy to meet during one of the research seminars, where he acted as a reviewer for my very first academic work. I learned a lot from your supportive, kind and constructive feedback then, and all of our many discussions that have followed have been so helpful for me as I made my way in the academic world. I am thankful to you also for inviting me to the **Digital Enlightenment Forum fireside chats**. These experiences have helped me to gain confidence in my ideas and research. Most of all, I am really grateful for the support you have provided throughout this process as a friend and academic mentor.

My warmest thanks go to my family, my husband **Indrek** and my children **Henri** and **Marii.** Henri and Marii, you have grown so much during this time, and you have been and always will be the biggest inspirations in my life. Let the culmination of this PhD process be a reminder to you to always believe in yourselves, surround yourselves with kind and supportive people and dream big. This experience would have been so different without constant support from **Indrek:** thank you for believing in me and supporting me in all of the steps I have chosen to take, which have at times meant shorter and longer trips away from home. I am also grateful to my mother **Maie**, my father **Endel** and his partner **Tiia,** as well as my sister **Laura** and her family. I wouldn't be who I am or where I am without your support.

1. INTRODUCTION

"We realize how much it is woven into our life when it fails.
Then the depth of our dependence on specific technology is realized."
Quan-Haase (2016:1)

Data, especially big data, have become a value, a promise and a hope for a better future. This value is related to the possibility of turning every human interaction and behaviour into data points that are tracked, collected and analysed, forming the basis of a datafied society (Mayer-Scöhnberger and Cukier, 2013). As more and more aspects of people's lives are mediated through digital tech giants (especially big tech companies from the US known by the acronym GAFAM[1] and from China known by the acronym BAT[2]), ever-increasing amounts of information about their consumption habits, social networks and locations are compiled. In media studies, this process is often called deep mediatisation (see Hepp, 2020a). Through owning, commercialising and managing digital technologies, multinational corporations and transnational organisations have become dominant governing actors alongside traditional governments (Suzor, 2019). For technology corporations and telecommunication companies, harvesting enormous amounts of user data has become the imperative business model. In doing so, researchers (e.g. Zuboff, 2019:17) argue that those companies accumulate vast domains of new knowledge *from us*, but not always *for us*. Other actors besides digital technology companies (e.g. governmental institutions) are also increasingly becoming involved in controlling personal data and producing value through different data governance models (Micheli et al., 2020:3) in the hopes of offering more efficient and better governance. This means that much of today's governance is increasingly reliant on (big) data.

Thus, data do not just naturally appear. They are always collected and manipulated through being made "algorithm ready" (Bucher, 2018:5), "scrubbed" (Gitelman and Jackson, 2013:7) and "cleaned" (Kennedy, 2016:108) for computer algorithms to use in their calculations. Hence, they are always shaped by human decisions related to selection, judgement, interpretation, action (Lindgren, 2019:2) and filters (Kennedy, 2016:110), and therefore data are *human artefacts* (Krippendorf, 2016), behind which there are assemblages of people, places, documents, practices and technologies which all make data products of complex processes in order to be useful for the contexts in which they appear (Ribes and Jackson, 2013, as cited in Svensson and Poveda Guillen, 2020:72). Technically, as Kitchin (2014:2) in his *Data Revolution* book argues, what we consider to be data are actually *capta,* i.e. "units of data that have been selected

[1] GAFAM refers to the American-based digital technology companies Alphabet-Google, Amazon, Facebook, Apple and Microsoft (for more, see Van Dijck, 2020)

[2] BAT refers to the digital technology companies Baidu, Alibaba and Tencent, based in China (for more, see Van Dijck, 2020)

and harvested from the sum of potential data". But as the term "data" is ingrained in our language, we use the term even when it may not be the most appropriate.

The term "big data" has, from the start, been related to the "widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible" (boyd and Crawford, 2012: 663). "Data-driven", often used in parallel with such concepts as "data-informed" and "data-based", algorithmic approaches are imagined to give us perfect information, real-time insights and smarter decision-making (Bonde Thylstrup et al., 2019), and therefore often debates related to mass data collection tend to present them as among the potentials for greater efficiency and (state) security.

Many of these beliefs referred to represent the dominant big data imaginaries. Imaginaries, often used in the context of examining scientific and technological progress (Jasanoff, 2015), are considered to be both factual and normative, meaning that they describe both how societies are thought to be and how they should be (Hockenhull and Cohn, 2021:303). These dominant imaginaries about envisioned data futures often involve not only imagining possible tomorrows, but they also affect the present development of data technologies and data arrangements in which they are embedded (Ruppert, 2018:4). Therefore, I will use "big data imaginaries" as an analytical concept through which I will analyse the ways in which data technologies are imagined by the pioneers whose understandings of the societal and economic outcomes of new data technologies I consider to have shaped data arrangements, affected the technologies that have developed and determined the ways these technologies are used. This will impact the ways datafied futures are experienced by individuals as data subjects, citizens and data producers or researchers as data consumers. In this thesis, imaginaries are thus considered the outcome of complex individual and collective sense-making activities connected with shared ideas about technology, which are often impacted by individual or collective ideals, ideas about technology, and fears about negative societal effects.

Thus, this thesis aims to analyse how dominant big data imaginaries are actualised and elaborated by data experts as data pioneers, and how this has affected scholarly practices. Imaginaries and the ways they are realised in practice also affect individual and institutional agency, based on commercial and value imperatives and technological affordances.

Big data is considered in this thesis mainly as a social construct which is needed to fulfil certain societal needs (Schäfer, 2011; Jasanoff and Kim, 2013; Van den Boomen, 2014; Quan-Haase, 2016). Thus, this thesis takes an approach that sees people as the central drivers of change. As a result of societal change and the corresponding societal needs, people develop new technologies which tend to lead to technological progress (Winner, 1993). Unlike technological determinism, technology is not considered in this thesis to be autonomous and self-directed (Winner, 1993). Rather, similarly to the approaches of Quan-Haase (2016) and Winner (1993), I consider how society or social groups attribute

meaning to technology and its uses or impacts. Therefore, technology is a social construct that receives its meaning and relevance from society. All technologies, as Thatcher and Dalton (2022) in their latest book emphasise, are socially and historically contingent, meaning that the material conditions under which they have been designed and used impact the meanings they produce in certain communities and societies. This also means that the public discourses and imaginaries about specific technologies should be analysed and described by taking into account the specific contexts in which they appear, i.e. in this thesis therefore it is important to acknowledge that the interviewed experts are from Estonia.

My interest in the big data imaginaries of Estonian data experts is derived from the fact that, in media (see Minevich, 2021) and branding web pages (e.g. e-Estonia.ee), Estonia has often been portrayed as a leader in digital technologies. In 2017, *Wired* magazine even called Estonia "the most advanced digital society in the world" (Hammersley, 2017). Nyman Metcalf (2019) has claimed that the aim of developing a citizen-centred public administration in Estonia has resulted from a culture of risk-taking, a highly educated population and an enthusiasm for new technologies. Algorithm Watch reports from 2019 and 2020 indicate that automated decision-making as a process of implementing and delegating tasks to digital systems (Kaun, 2021) has been increasingly introduced in many public sector institutions in Nordic countries (e.g. Broomfield and Reutter, 2021; Jørgensen, 2021; Kaun, 2021), as well as in Estonia (Männiste, 2020). Although public-private partnerships have been successful in developing e-governance frameworks, it is not clear how the emergence of big data and other data technologies has challenged data experts in Estonia. Several examples, including the E-Residency[3] project and Estonian experts' involvement in developing distributed digital infrastructure providing health solutions during the COVID-19 pandemic (Vihma, 2020), indicate a future-oriented vision of data-driven governance. Developing these solutions often includes offering counter-imaginaries, i.e. narratives, tools and practices actors employ when seeking to counter the threats of datafication (Kazansky and Milan, 2021), to the dominant imaginaries about datafied futures. Based on this, I consider experts in this thesis data pioneers who form pioneer communities (Hepp, 2016; 2020a; **STUDY III)** whose imaginaries about datafied futures may define the ways they materialise.

The ways that possibilities in relation to big data are imagined by those most responsible for building a digital society – data experts – have not been studied and are hidden from the general public. Previous studies have concentrated on the practices of experts from the public sector (Redden, 2018; Jørgensen, 2021). This cover article will take a wider approach in discussing the ways datafied

[3] E-residency, or "virtual residency", is a project initiated by the Estonian government to provide entrepreneurs and residents of other countries global access to Estonian e-services via state-issued digital identity in the form of electronic smart cards (eID) (Tammpuu and Masso, 2018).

futures are imagined and practised by Estonian data experts, as well as how they have affected scholarly practices **(STUDY II)** and researchers as data experts, by whom, as Van Dijck (2014: 198) claims, datafication has been seen as a "revolutionary research opportunity to investigate human conduct." Therefore, pioneers in the context of this thesis are not just the individual data experts who form pioneer communities (Hepp, 2016; Hepp, 2020a; Hepp, 2020b), but also Estonia as a digitally advanced society and scholars whose research is based on big sets of social media.

The algorithmic processing of data, which has opened up endless new opportunities, also challenges the protection of data privacy, as one of the core values in datafied societies, both regarding private companies and states, especially as the possible future uses of data collected by different entities can hardly be foreseen (Mayer-Schönberger and Cukier, 2013). Although the European General Data Protection Regulation (European Commission, 2016) was enacted to overcome the possible vagueness of potential uses, in the case of an unforeseen "legitimate interest" on the part of the processor or a third party data can easily be used for any purpose (Roßnagel, 2020:6). Authors (e.g. Franzke, Muis and Schäfer, 2021) have also emphasised that not all of the concerns regarding privacy that arise from different government data projects cannot be handled by strictly applying privacy laws or data management regulations. They propose that data projects should also be subject to ethical frameworks, which would help in ensuring ethical practices and accountability. But they also emphasise that this handling should include a dialogic review or deliberation process (for example, DEDA, developed by the Utrecht Data School; for more information, see Schäfer and Clausen, 2021) because otherwise ethical frameworks may end up being just posters, guidelines or manuals that do little to help practitioners.

From data subjects' viewpoint, the algorithmic processing of data also means that users often have no way of knowing what is done with their data, how algorithms work or what others can do with the data (Eslami et al., 2015; Woodruff et al., 2018; Carmi, 2020). Thus, individuals need to put a great deal of trust in the creators of algorithms and the providers of these services. This also means that the perception of institutional trust, i.e. trust individuals have in different institutions that track, collect and analyse their data through various data points, has become increasingly important **(STUDY I)**.

In this cover article, I especially concentrate on data experts who are in this thesis considered to be both Estonian experts working with data or developing data-based solutions in relation to mobilities (**STUDIES III and IV**) and researchers who have concentrated on analysing social media big data (**STUDY II**). Through these specific cases, I will analyse the imaginaries specific actors have and how these imaginaries may impact practices and understandings about data on various levels. This cover article emphasises, therefore, the role people have in this process of either collecting, processing and analysing data, or using data technologies. Too often, as Cohn (2019:92) argues, "humans are distanced or forgotten in this process which makes the algorithms they create appear far more like black boxes than they actually are". Those technologies are, however,

mundane and profoundly non-magical, although they are often overly mystified. Big data imaginaries of desired and possible futures are performed through the relations between people and technologies (Ruppert, 2018:13). Through concentrating on those relations, actors are considered variously agentic in what imaginaries develop. This means that not all actors have similar capacities to produce collective imaginaries (Jasanoff, 2015), and not all imaginaries will materialise. The materialisation of imaginaries has also been affected by the roles of specific actors, e.g. data producers, consumers or distributors, and which internal and external barriers are imposed in relation to their roles in datafied societies. Imaginaries have also constituted useful frameworks in previous research which has analysed the values and visions guiding European policymakers (Rieder, 2018), as well as the data imaginaries in the larger states of the Global South (Brazil, India and China) (Mahrenbach and Mayer, 2020). Academic work has also explored the ways that imaginaries of big data affect expert knowledge and the epistemic authority of state institutions producing specific data futures through changes in practices of data production and analysis (Ruppert, 2018). Additionally, authors (e.g. Lehtiniemi and Ruckenstein, 2019) have pointed out alternative imaginaries which describe how data activists or civil society actors counter the evolving threats of datafication.

This thesis will contribute to previous research by analysing which kinds of big data imaginaries are elaborated and actualised by Estonian data experts in private, public and third sector institutions **(STUDIES III and IV),** as well as how the emergence of (social media) big data has changed scholarly practices and knowledge creation **(STUDY II)**. Experts' visions, as well as the ways possible futures are realised in everyday practices, will also help to clarify the challenges those data futures pose to both individual and collective agency (**STUDY I**). Datafication, e.g. the quantification of every aspect, plays a central role in configuring social relations (Van Dijck et al., 2018: 36) and therefore has and will have a big impact on understanding and practices at every level. To better understand current practices, there is a need to understand if and how the emergence and the access to big sets of data (i.e. social media data) have not just challenged data experts, but also how big data has been perceived and what effects it has had on scholarly practices.

In this cover article, I will concentrate on how the norms, ideals, commercial imperatives and technological affordances of digital environments have challenged the processes of data technologies (from data collection to final output) and both individual and collective agency. This cover article engages empirically with the implementation of datafication practices and will highlight the imaginaries and practices, as well as challenges, that emerge when those systems are implemented.

This cover article is based on four empirical articles conducted between 2016 and 2020. The first part of this doctoral study (**STUDY II**) focused on exploring the methodological changes expressed in research practices. For this purpose, a systematic literature review method was used, and 120 peer-reviewed empirical articles that used social media big data were both quantitatively and

qualitatively analysed. **STUDY III** and **STUDY IV** focus on Estonian data experts (n=24) as a pioneer community and their understandings of social datafication, algorithmic governance ideals and practices. It is based on 24 in-depth interviews with Estonian data experts working with human mobility data, i.e. regulating everyday mobilities, forced migration or digital migration (for example the E-Residency programme). Experts working with mobility data were chosen, as in this field data-based decision-making is increasingly used in mobility management. The third part of the thesis (**STUDY I**) focuses on explaining peoples' concerns regarding privacy and the relationship between trust and privacy concerns. Although the importance of trust has been examined in previous research, this particular study makes a clear connection between institutional privacy concerns and institutional trust. For this purpose, data from an Estonian representative population survey of 1503 respondents were quantitatively analysed.

This cover article is structured as follows: the first part gives an overview of relevant theoretical concepts, such as big data imaginaries and agency in datafied societies; aspects of the specific Estonian context are also briefly introduced. The second section describes the methodological part of the studies. In the third section, I present the main results of my four studies and, in the following discussion chapter, I explain and evaluate the main results, show how they relate to previous research and propose a model through which the triple roles different agents have in datafied society can be discussed. The cover article ends with a conclusion and supplementary summary in Estonian.

2. THEORETICAL FRAMEWORK

From the start, there has been attempts to define what big data is. Often, the definitions rely on the "3 V's": volume, value and veracity (see Laney, 2012; Andrejevic, 2014; McNeely and Hahm, 2014; Schroeder, 2014; Mittelstadt and Floridi, 2015), dimensions which tend to mostly focus on data properties (Rieder and Simon, 2016). Others have viewed big data as having profoundly changed how people think, work and live (see Mayer-Schönberger and Cukier, 2013). Therefore, the precise definition for big data remains elusive, and it can be viewed very differently by different stakeholders. The vagueness of big data also offers a convenient way to use it as an "umbrella term", broad enough to be applicable to almost anything technology-related (Rieder, 2018). Although technology plays an important role in making vast amounts of data collection and analysis possible, the "power of big data" relies on how that data and their potential are imagined (Beer, 2016: 9) by different actors. Therefore, as the creation and use of new technologies have produced new social imaginaries (Mansell, 2012), a similar thing can be said about the "big data revolution" (Kitchin, 2014), of which different actors have provided different possible futures.

The next chapter will aim to contextualise the main big data narratives and visions prominent in media and scholarly work. This approach will help in pointing out specific economic, institutional and epistemic challenges which contribute to the production and perpetuation of big data imaginaries and which will be helpful in discussing changes in power, agency and control (Hintz et al., 2019) in datafied societies.

2.1. The big data imaginaries

The concept of "sociotechnical imaginaries" has been widely used in research as an analytical tool to describe and explain technoscientific projects, social constellations and politics. Jasanoff (2015:4) describes sociotechnical imaginaries as "collectively held, institutionally stabilized and publicly performed visions of a desirable future" that are animated by shared understandings of forms of social life and social order and made attainable through the design of technological projects. According to Williamson (2015: 2), such futures are "produced by particular social groups within specific social contexts, and they are also projected through the design of particular kinds of technologies to express a view of particular futures in which those kinds of technologies are imagined to be integral embedded parts". In terms of datafication and data technologies, this provides a good way to account for how data and technology are embedded in the social context (Jasanoff, 2015: 2–3), and therefore helps to account for the interplay between the designs of technologies and the social arrangements that inspire and sustain data production.

As Rieder (2018:103) argues, visions of possible futures are always "culturally and temporally particular, embedded within specific socio-political environments," which means that it is important to account for the specific contexts in which those visions emerge. The contexts may end up defining the ways in which those possible futures can be achieved. This is best seen in the ways surveillance is approached and justified in China and in Western countries. Similarly, when discussing challenges and opportunities datafication poses in Europe, especially in Nordic countries, it has to be taken into account that the imaginaries of the welfare state and how citizens' agency is understood in those contexts inevitably impact future visions of the use of data technologies in particular regions.

The initial understanding of sociotechnical imaginaries focused on how nation-states, governmental actors and public institutions envision and enact technoscientific developments has recently also become an object of criticism. Research (e.g. Olbrich and Witjes, 2016; Mager, 2018; Lehtiniemi and Ruckenstein, 2019; Felt and Öchsner, 2019) has highlighted how imaginaries and counter-imaginaries (Kazansky and Milan, 2021) are also articulated and enacted by corporate actors, civil society, research communities and other organised groups in processes much more complex and non-linear than envisaged in the original concept.

Jasanoff and Kim (2009) emphasise that when certain imaginaries are widely accepted and used, they may shape trajectories of research and innovation, steering technological progress and public/private expenditure. The real force of those imaginaries, however, becomes evident in practices that cultivate different forms of cultural capital, in the case of big data in calls to change ways of thinking about data production and the uses of data technologies. Therefore, it is important to consider the "data practices" that generate digital data and to acknowledge that the ways these data are interpreted and made meaningful also generate particular effects and social implications, since data and algorithms that process them are consequential to "what is known" and can influence decision-making and other activities" (Ruppert, 2015). Throughout the thesis, I will treat data practices as social and technical instantiations of particular future visions, as practices that operationalise imaginaries in the present, often in ways that are messier in practice than was anticipated and desired.

Most of the imaginaries related to big data are supported by the widespread belief in today's society that data are resources that should be utilised for economic outputs. In those understandings, data are sometimes conceptualised through different metaphors, which in their own ways contribute to the ways data are understood on different levels. Puschmann and Burgess (2014) refer to the claim that "big data is a force of nature to be controlled," which is associated with the natural force of water. According to this approach, society is drowning and will have to deal with data tsunamis in different ways. Puschmann and Burgess (2014) argue that the analogy of water fits in the sense that water is neutral and able to exist without humans. With the appropriate technology, however, water and data can be harnessed. Big data is also referred to as

the "new oil" (Helbing, 2015) or the "gold mine" (Rieder, 2018). As mentioned above, this cover article opposes, as do many other scholars (Andrejevic, 2014; Bowker, 2014; Kitchin, 2014; Thatcher, 2014), the idea of data being something neutral. I consider concepts of currency (e.g. Thimm et al., 2017) or capital (e.g. Sadowski, 2019), which refer to the ways that everyday actions are turned into hard-money value, more suitable. These concepts best reflect big data as strategic resources and advantages for organisations. Thus, first and foremost, big data are imagined in dominant imaginaries as a game-changer, the hope for higher profits, efficiency and bright future prospects **(STUDIES III and IV).**

Besides big data being seen as a profitable resource, discourses surrounding big data and connected technologies often include the belief that making something more data-based or data-driven will lead to more objective and reasoned decisions (McAfee et al., 2012). This rhetoric, as Cohn (2019) argues, is premised on the belief that the sheer amount of data that algorithmic output is based on will create an undeniable connection to objective reality. This "digital positivism" (Mosco, 2015) promotes the epistemological assumption that we can technologically control big data collection and analysis and do it to the extent that data "will speak for itself" and become inherently meaningful**.** This relates to the assumption that there is a "self-evident relationship between data and people, where interpreted aggregated data is seen to predict human behavior" (Dencik, 2020). But, as Safia Umoja Noble (2018:1) emphasises, data and algorithms are not neutral or objective; they are "anything but." Digital technologies are designed by humans, many of whom "openly promote racism, sexism, and false notions of meritocracy" (Noble, 2018: 2), and this means that norms, values and assumptions about gender, race and sexuality "are encoded in and reproduced through the design of socio-technical systems" (Costanza-Chock, 2020:4).

This understanding also relies on a particular epistemic viewpoint in which there is only what appears in the data. In these cases, the statistical maxim that "correlation does not imply causation" is abandoned for the maxim that "numbers speak for themselves" (Andrejevic, 2014; Bowker, 2014; Kitchin, 2014; Thatcher, 2014). As Siegel (2013:90) claims, we do not need to worry about causation, explaining the why, when the objective is to predict the world rather than to understand it. Computational social science, which emerged in relation to opportunities to, for example, access vast amounts of social media data and use it for research purposes, first generated the main innovations in relation to big data analysis related to data analysis techniques and tools (see He et al., 2015; Park et al., 2015), i.e. in predicting the world. Although computational techniques applied to digital data are seen as solutions to the data turn in social sciences (see Hindman, 2015; Keuschnigg et al., 2018), some authors (e.g. Thatcher et al., 2016; Resnyansky, 2019) emphasise the asymmetrical power relations that have a significant influence on the ways that social science is practised (**STUDY II)**. Moreover, recent research (e.g. Van Es et al., 2021) highlights the importance of a critical, reflective attitude towards the tools, i.e. software applications, used for research purposes, and aspects of digital

methods. Van Es et al. (2021) note that tool criticism is especially needed as some of the tools social studies use are imported from other institutional contexts and can be developed for other purposes. There are also freely accessible and easily usable tools developed by the Digital Methods Initiative in Amsterdam, which, as Van Es et al. (2021) emphasise, are often used without acknowledging their possible limitations. Therefore, it is increasingly important to understand and evaluate both the opportunities and the limitations of the use of these tools by scholars.

In seeing datafication as closely related to both commercial profits and an overall increase in efficiency, in Western societies it is also often considered an inevitable part of innovation, as were many older technologies. This has been portrayed through different narratives, each of which supports the overall understanding that data collection today has become inevitable and even necessary, e.g. for security reasons. Thus, the fact that data form an object of economic hope and a mode for solving complex social problems has encouraged states to innovate and disrupt traditional modes of governance. Data-driven and predictive technologies are seen to provide ways of being able to do more, better, faster and more cheaply through automation or augmentation (Maciejewski, 2017; Klievink et al., 2017). Moreover, this data-driven approach is often seen as a way to respond to the growing complexity of society and, in discussions of computational social science, this is a shared imaginary of both researchers and practitioners. However, recent research in Denmark (Jørgensen, 2021) and Norway (Broomfield & Reutter, 2021) on data-driven approaches implemented in public administration emphasise the risk of widening the power asymmetries between the state and citizens through these optimistic narratives of data-driven governance, as they also reflect a simplistic view of technology as a tool that must be provided, accessed and learned rather than something that mediates relationships and reconfigures power (Van Dijck, 2014) and structures of inequality (Eubanks, 2018). As Broomfield and Reutter's (2021) research shows, although practitioners view the use of "technology for the good of society," actually effectiveness is the dominant variable guiding data-driven approaches. Similar claims were asserted in **STUDIES III and IV**, which illustrate that, for many practitioners in Estonia, those data-driven practices are often justified as pursuits of "the social good," although the meaning of this term remains vague in datafied societies. Moreover, as the recent article by Magalhães and Couldry (2021) shows, this approach to collecting and analysing big data, or using AI for the social good, is increasingly used by the big technology companies: Facebook, Google, Amazon, etc. They (Magalhães and Couldry, 2021: 354) emphasise that not only are these companies using datafication for the social good, but the practice also has a hidden and rather consequential goal: "progressive reconfiguration of the social domain itself in ways that position those technology companies as privileged providers of social solutions and privileged purveyors of social knowledge." Magalhães and Couldry (2021:354), therefore, argue that the "social good is not a neutral fact

but rather a set of socially constructed parameters by reference to which good and consequential actions in the territories we share are evaluated".

Thus, while digital governance is often promoted as an objective means to provide more efficient and targeted services, critics view the datafied turn in government decision-making and services as highly political and one that undermines the rights of the already marginalised while exacerbating inequality and discrimination (Eubanks, 2018; Alston, 2019; Benjamin, 2019). Researchers (e.g. Angwin et al., 2016; Eubanks, 2018; Dencik et al., 2018) have also raised concerns about how the logic and systems used to rank and score people for commercial purposes can now also be found within the public sector, influencing decisions about funding, resources and front-line services (see Algorithm Watch, 2019 report). Redden et al., (2020) believe that, as too little is known about where and how changes are taking place, the larger political implications of the data systems introduced and their economic underpinnings may also remain unknown. Thus, the lack of information publicly available about data technologies makes it often nearly impossible for publics to know how data systems are developed, implemented and used, which limits public debate and civil society involvement. Therefore, too often, researchers' and civil society actors' role is diminished to pointing out the problems of already implemented systems, although they could be actively involved in developing solutions (Kasapoglu et al., 2021).

Additionally, research on the Norwegian public sector has shown that innovation happens at the sectoral and organisational levels (Difi, 2018 as cited in Broomfield and Reutter, 2021:80). This poses a risk that there is no clear overview of which kinds of data projects are happening on a wider level and, as Brauneis and Goodman (2018) argue, this may pose a major transparency problem for public sector institutions. Furthermore, how can we evaluate or even criticise any of the data-driven solutions if we do not have an overview of where they are deployed, especially in public sector institutions?

Data-driven solutions, especially algorithms, are therefore often overly mystified. Discussions also suggest opening up the "black box" of algorithms (see, e.g., Pasqual, 2015). Through this, it is hoped that problematic automated decision-making or algorithms could be made accountable and help individuals gain back control over their privacy, understood here as personal control over information and access (Solove, 2008). Discussions surrounding the "black-boxed" nature of algorithms clearly illustrate the changing nature of data relations. Data rely on the technical procedures of governing institutions that produce the data, including different private companies, state agencies and administrations. Governing here means that these institutions are in a powerful position to define the character and structure of data and metadata, as well as their possible uses (Breiter and Hepp, 2018:391). This results in a new form of power exercised mostly by "those with access to databases, processing power, and data mining expertise" (Andrejevic, 2014:1676). New questions about what and whose power is exercised through such practices, and to what degree such

exercises of power are satisfactorily held accountable (Couldry and Powell, 2014:2), are the focus of many discussions.

Thus, the use of bigger sets of data also creates novel understandings of what visibility and exposure are, as well as creating a new type of asymmetric relationship between data subjects and the institutions and organisations that utilise big data. Tene and Polonetsky (2013) suggest that data privacy protections move from a concern with the point of collection to a concern with transparency, access and accuracy. These new power dynamics, however, create imaginaries in which concerns about privacy keep centring more on the collection of data and less on the use of data. Whilst data collection can be regulated through legislation to protect individuals from some of the consequences, this does not change the basic mechanisms and laws which give power, knowledge and wealth to the Big Tech companies, as Tene and Polentsky (2013) have emphasised. Moreover, as Broomfield and Reutter (2021) have argued, privacy and legal issues are often equated with trust, which means that, at least on the public sector level, if privacy is protected and current regulations are adhered to, there is little to be concerned about. Research from **STUDY I** supports the claim that concerns related to privacy are related to institutional trust, but the constantly changing nature of the datafied society may today produce totally different results. During the six years after the **STUDY I** data was collected, the practices implemented by both private and public sector institutions became more data-driven, and the possible societal consequences have been further discussed in the media. The lack of transparency in the use and development of data technologies also raises the question: how can you worry about something you do not know about? This concern is even more emphasised by the increasing use of automated decision-making systems (see Algorithm Watch, & Bertelsmann Stiftung, 2020), as AI systems that rely on deep learning processes can run independently of human control (Zerilli et al., 2018). Therefore, it is unknown how such processes reach decisions.

Many of the previously described changes have led to imaginaries and practices which also highlight the importance of the openness of data (and practices surrounding data technologies). Several of these changes also actively seek solutions for the emerging problem of data ownership, for example through such models as data sharing pools, data cooperatives, public data trusts and personal data sovereignty (Micheli et al., 2020). Those models not only value the openness of data (e.g. public data trusts, such as Open Data Institute) but can also promote a different and fairer data economy, where data subjects are imagined as key stakeholders together with digital service providers (e.g. personal data sovereignty). All of those models propose different kinds of approaches to data, where the data owner and data subject can be considered the same or different, and they therefore may have totally different agentic roles. This implies that the far-reaching consequences of datafication also challenge the agency of both individual and institutional actors.

2.2. Agency challenged by big data imaginaries

As many of the previously described dominant big data imaginaries involve certain tasks increasingly being assigned to technology, researchers have started to question whether deep mediatisation, i.e. an advanced stage of the process in which all elements of our social world are intricately related to digital media and their underlying structure, enhances or limits agency, and on which individuals or institutions it has such effects in the construction of the social world (Couldry and Hepp, 2017). Although there are different conceptions of agency, I will base my understanding of agency on the work of Nick Couldry (2014: 891), who defines it as "the longer processes of action based on reflection**,** giving an account of what one has done, even more basically, making sense of the world so as to act within it." According to him, agency is, therefore, a reflective practice. Couldry's approach to agency is similar to Layder's (2006), who argues that social analysis has to take into account the meaning that the social world has for the individual based on how the person understands and responds to their lived experience. Moreover, he argues that the way people construct their social existence helps them formulate their plans and intentions, and choices about the direction in which their lives should go are based on their experience (Layder, 2006). Thus, similarly to Couldry (2014), for Layder individuals are intentional, self-reflective and capable of making a difference in the world.

Emirbayer's and Mische's (1998) approach provides another important aspect to consider when discussing human agency. In their definition, they refer to three elements, each connecting agency to the past, present and future. These "constitutive elements" that they talk about in relation to human agency are **iteration, projectivity and practical evaluation,** which all indicate that agency depends on **routinised practices, goal-seeking and purposive activities** (Emirbayer and Mische, 1998). However, in the past ten years, as a part of the emergence of a new wave of datafication, Couldry & Hepp (2017:145) argue that the basic conditions for being a social actor have changed, as in many societies social actors are expected to be available for interaction through digital platforms. The social and cultural transformations of datafication have led to new types of reflective agency, in which attention should be paid to the ways different actors respond to processes of data collection and analysis, and how they use data to "meet their own needs" (Couldry, 2014:892). However, as Kennedy and Moss (2015) argue, for publics to have greater agency and reflectivity, certain changes have to emerge in data practices.

With the new wave of datafication, questions about algorithms as agents have also emerged. As Goffey (2008:18) argues, algorithms may be agents as they "do things". But their agency clearly stems from the socio-technical assemblages in which they are embedded (Introna, 2016:20). Therefore, as I will also later discuss, although algorithms do shape social and cultural formations and can directly impact individual lives (Beer, 2009: 994), they are not autonomous (at least, not yet). Rather, their agency is interlinked with that of the people who write them and run them, and with the users who interact with them.

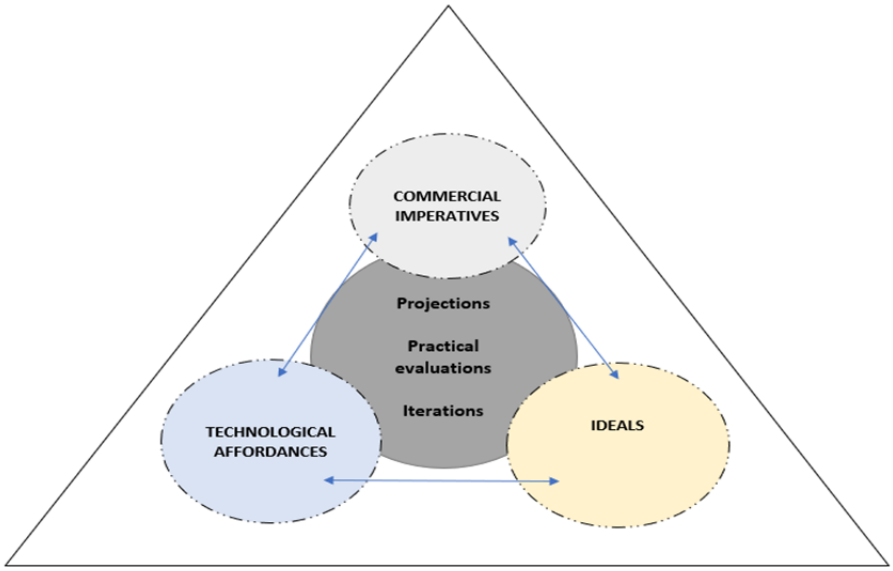

Figure 1. Three interrelated elements predominantly affecting agency and the materialisation of big data imaginaries. The arrows represent the interconnectedness of the elements. The model is based on the Klinger and Svensson model (2018:4663).

For the purpose of discussing the changes in data relations and challenges datafication poses on both individual and collective actors, I have adopted the Klinger and Svensson (2018) model, in which they discuss agency in terms of media logic. Their model (Klinger and Svensson, 2018) discusses agency in terms of **iteration, projections and practical evaluation,** which are accounted for through the three dimensions of **content production, distribution of information and media use,** consisting of three elements: **underlying ideals, commercial imperatives and technological affordances**. I have operationalised their model in terms of these three elements and in terms of iteration, projection and practical evaluation. Although all of these elements have different weights when certain big data imaginaries are materialised by specific actors, I see them as interrelated in practice and dependent on the particular roles actors in these specific moments may have. In discussing data agency, here understood as agents' capacity to act intentionally in relation to personal data and its collection and use by different actors, I consider individuals and institutions in datafied societies to have three main roles: **data producers, data distributors and data consumers**.[4] The ways in which certain elements affect practices and experien-

[4] The terms "data producer" and "consumer" are in this context used similarly to the way they are used by Meijas and Couldry (2019: 269–280), who argue that consumption in the era of digital platforms consitutes a new form of exploitive capitalist relations. They point out how today's relations are centred around data. Data is not simply abstracted from us, but often obtained through different kinds of relations which, as they argue, we may think we

ces on the individual level can, as indicated in **STUDY IV,** affect the ways in which data futures can be imagined on the collective level as data experts.

2.2.1. Commercial imperatives

The first element of agency considers today's commercial imperatives, which are centred around humans actively and intentionally spending time on communication platforms, as Klinger and Svensson (2018) argue. Through platforms, users leave the "footprints" of digital media use that Breiter and Hepp (2018: 387) describe as digital traces. These digital traces form the data which are the basis of all future processes. As what data users leave on those platforms, either intentionally or unintentionally, becomes the main resource fuelling data technologies, the user also becomes a commodity. Most of the data on people are not public assets and are increasingly privately funded, collected and analysed (Van Dijck, 2020a; Couldry and Yu, 2018), e.g., through social media platforms, and therefore are also not openly accessible to most other parties. By algorithmically mining these data, different private companies can surveil users with commercial intent, for example in order to predict a user's future buying patterns or to personalise services. Often, as Bolin and Andersson Schwarz (2015: 5) argue, the "socially explainable 'who' behind this (behavioral) pattern is less important than the algorithmically predictable behavioral 'how.'" This means that people are not governed in relation to their individuality, as Ruppert (2011:218) emphasises, but as members of populations.

Those projections about users' future activities or behaviour both offer new opportunities and challenge users' agency. Greater access to different kinds of "locative" services and news or ways to find nearby friends through different apps are just a few examples Future patterns are mined through algorithmic processes that tend to be opaque not just for users but also for the developers themselves. In this way, data "produced" by users travel through different infrastructural solutions and are used in ways that are often not related to the initial purpose of data collection (see Magalhaes and Couldry, 2021). In the digital advertising and marketing industry, where social media play an important role, it has been common practice to source (or access) data through different partnerships and to use data for different purposes than initially intended (Jarvenpaa and Markus, 2020). These, and several other kinds of partnerships, illustrate how data are considered a strategic asset for many firms (Spiekermann, 2019). Partners, as powerful industry players, have their own interests and business models (Braun, 2013:127). For users as consumers, the other players, their bargaining power in these data relations is often relatively non-existent. The individuals and organisations which create or provide platforms

---

have retrospectively consented to. Furthermore, considering that the data we produce can also produce value for ourselves, we also end up being both the authors and consumers through these relations. Therefore, these concepts are useful here to described increasingly commodified data relations.

with complementary tools, products, or services for specific platforms (Gawer and Cusumano, 2014) include app developers, businesses and partners, advertisers and marketers, content creators, i.e. influencers, and media publishers. Through their practices, these different parties form an interconnected platform infrastructure that connects social media and the partners (Van der Vlist and Helmond, 2021). This, however, means that power is dispersed and exercised through infrastructure in which the gateway function of APIs (Application Programming Interfaces) is an important source for the "infrastructural power" held by platforms (Van Dijck et al., 2019). Through these "relationship advantages" (Broughton, Micova and Jacques, 2020), platforms increase their power, as well as the "opacity bias" (i.e. a lack of transparency in relation to programmatic advertising).

There are several other ways in which algorithms as data technologies are considered to be opaque and black-boxed (Pasquale, 2015). First, algorithms typically involve intentional secrecy, as data and codes are protected by companies and administrators as valuable intellectual property (Burrell, 2016). This means that observers do not have access to algorithms because companies do not make them public. At the same time, as Prainsack (2019) notes, big tech companies, through these platform infrastructures, increasingly act as key social and economic intermediaries, often providing essential services, products and infrastructures in exchange for people's personal data. This has been especially apparent during the COVID-19 crisis, which has forced people around the world to work from home. The interconnectedness between different platforms and partner services makes it hard to find suitable alternatives to Big Tech services (e.g. Teams and Zoom).

But it is not just private companies whose practices should be considered here. As I will discuss in more detail in Chapter 4.3, public and third sector institutions also share this understanding of data as a resource (**STUDY III**). Their approach to data is not to commercialise the value produced through data technologies but to offer new value through data-based or data-informed (**STUDY III**) analyses and data technologies to different stakeholders. Unlike private companies, in the case of public sector institutions, individuals may not have any say in whether or not to use certain services, even if there are problems. Therefore, as Barassi (2019: 419) has stated, citizens as data subjects are "systematically coerced to digitally participate to provide data".

In cases where companies decide to share their algorithms with users and researchers, another dimension of opacity emerges: technical illiteracy. As algorithms are made of code written in programming languages, it is hard for a user without specific training and knowledge to understand them. Moreover, in the case of researchers **(STUDY II)**, this leads to new digital divides where specific training provides some with an advantage in understanding the inner logic of algorithms and eases access to data. If we consider machine learning, training in programming languages doesn't even help because, as previously mentioned, algorithms evolve over time in ways that are typically unintelligible to humans (Burrell, 2016). Moreover, the sheer size of algorithmic systems makes it im-

possible to understand them. In many cases, specific services may rely on billions of lines of code (Metz, 2015), and this makes it impossible for anyone to actually identify which part of the system is responsible for a specific decision. These digital divides, however, are not just technical: even one's native language is increasingly a disadvantage in using data technologies and doing analyses. One possible way to at least partly resolve these digital divides is the previously mentioned interdisciplinary centres and digital methods workshops/ schools, which offer access to not just tools but often to data sources that may otherwise be inaccessible. For individuals, in recent years several countries have used the public and free course "Elements of AI" which has been translated into several different languages (see Algorithm Watch, and Bertelsmann Stiftung, 2020).

Moreover, algorithms may draw on historical data shaped by long histories of inequality and discrimination, and thus algorithms are often considered to be biased (Barocas and Selbst, 2016). This also increases problems with inherent opacity. As several authors have emphasised (Crawford and Schultz, 2014; O'Neill, 2016; Eubanks, 2018; Chesterman, 2020), if those biased decisions are not dealt with, they could pose a serious threat to the notion of due process in democratic societies. This is also true of how past patterns are constantly repeated in new environments. These iterations also happen in relation to the partnerships described above. The relationship advantages that other institutional parties can get from increasing data collection support the further collection of data.

2.2.2. Technological affordances

Another important element in Klinger and Svensson's (2018) model concerns technological affordances and their impacts on agency. Nagy and Neff (2015) argue that the term "affordance", as it is commonly defined in the communication literature, does not capture the complexity of the interaction of people in the process of being afforded action within sociotechnical systems. According to Neff and Nagy (2016), people no longer treat or view smart agents as mere tools. Therefore such objects may be considered to have a technical agency that has unique participation status in interaction (Kummerheuer, 2015). Nagy and Neff (2015:1) have emphasised "that rather than thinking of technologies as having fixed capacities that are recognized by their human partners, imagined affordances allow us to describe user perceptions, attitudes, and expectations; the materiality and functionality of technologies and the intentions and perceptions of designers." Therefore, affordances should never be seen just as products of designed features or the practices of users but as being both designed by and evolved from use. As Hutchby (2001) has argued, this means that the features of the environment provide a framework for the possibilities of agentic action in relation to a particular object but they certainly do not determine them. Rather, they set specific conditions for how affordances are experienced. I propose that the ideals and norms of developers, as well as different imperatives of both

individuals and collectives and the actual quality of material, affect technological affordances **(STUDY III)**.

Individuals are affected by projections and expectations related to specific technologies (Klinger and Svensson, 2018): the technological affordances provided by the environment give them clues about what they are supposed to and can do. Through the process of algorithmic curation and representation, the formation and cultivation of publics are affected. Gillespie (2014) has argued that through opaque curation logic, algorithms produce "calculated publics", i.e. publics that do not exist before they are calculated, and only platforms know exactly what membership in a specific public entails. The ways in which certain content, users or paths are highlighted leads to sorting users into specific categories (Bodó, 2019). These categories, usually not transparent to the user, make them part of specific publics invoked by algorithms, for example "customers like you," which, as Gillespie (2014) notes, may have nothing to do with the publics users themselves would have sought out. Although Gillespie (2014) uses this as an example of how any user agency-specific publics can be formed, recent research has highlighted different ways that users employ different coping tactics and experience algorithms both positively and negatively (Pink et al., 2018; Hartley and Schwartz, 2020; Ytre-Arne and Moe, 2021). This means that through the specific affordances of a digital environment and users' understanding of how specific algorithms work, they may intentionally "play" with algorithms and, through this, disrupt predictable patterns. Moreover, as Lomborg and Kapsch (2020: 758) have noted: "While users may seem stripped of the agency once they are looped in algorithmic systems in everyday life, the small acts of actively curating, withholding or flagging information to tweak the system to enhance privacy and evade precise profiling are indeed subversive means to speak back to the system."

Technological affordances also shape how certain concerns are approached and perceived by individuals. Privacy, seen as one of the core values in a datafied society, is increasingly impacted by the dominance of big technology companies. According to Rössler (2015:23), something is only private if the individual can control access to it. This assumes data subjects possess what they are claiming control of. However, the current solutions for reclaiming this control tend to be merely iterations of past solutions. Moreover, some of the current solutions, e.g. "notice and consent" (where data subjects have to agree before using certain solutions or sharing their data), distract attention from the responsibilities of platforms (and other institutions) (Schmitz, 2014:1463). But, in reality, in a datafied society, the actual choice for the individual in many of those situations is to be excluded from social participation or from sharing their data. Although individuals increasingly have no choice in their exclusion from those processes, both with commercial entities and with wider processes for different reasons, research (Selwyn and Pangrazio, 2018) also indicates that for younger people sharing data with third parties may be an acceptable trade-off as they see more value in being active and sharing information on platforms.

There are also problems with relating to this type of privacy, mostly in terms of individual autonomy. The digital environment and datafication complicate it and raise the question: given that information extracted from individuals can reveal private information about entire groups or communities, shouldn't we be talking more about safeguards necessary for groups or communities who are exposed through the practices of the individual or institutions?

Klinger and Svensson (2018) have emphasised that affordances always result from interactions between two or more types of human actors and the technological structure between them. Therefore, practical evaluations of how best to use the system, user input and other complex situations shape the ways in which technologies are not just used but also how they are developed. Often this may leave users in situations where they can only adapt to the affordances posed by environments or platforms when designers choose the interests and values and how they are translated into the code of the software. For example, McKelvey (2014:589) has noted how, in recommender systems, different associations or relationships between users and users' interests can have very troubling or unexpected connotations more broadly. The specific tags used to make information, apps etc. findable, the placement of apps in the Google Play Store (e.g. a gay men's dating app next to a sex offender search app; for more, see Ananny, 2011) can end up associating two topics and, by recommending these to users, can be socially, ethically and politically problematic. As McKelvey (2014) has noted, algorithms involved in these processes don't distinguish among data they have been instructed to analyse and manipulate in terms of explicit political and social values. This, as Birhane (2021) has noted, also means that any data scientist (and other kinds of data experts working with data) involved in automating is also engaged in making moral and ethical decisions, and is not merely dealing with technical practices impacting individuals.

Collective actors have active roles in these asymmetric data relations as their ideals and practices result in specific affordances for users, and the affordances either foster dominant big data imaginaries or offer particular new counter-imaginaries.

2.2.3. Agency of pioneer communities: the driving force of datafication

The literature (e.g. Redden, 2018) has emphasised the significant role of active agents in using and designing data technologies. This cover article considers Estonian data experts and social media big data researchers active agents whose practices and ideals impact both the design and use of future data technologies. The term "pioneer communities" has been used (e.g. in Hepp, 2016; 2020a) in describing those agents, as have the terms "information commons" and "data commons" (Prainsack, 2019; Taylor and Purtova, 2019). Although Hepp (2020a) has used the notion "pioneer community" to specifically describe influential groups of the Quantified Self movement, Maker movement and Hacker movement, in this cover article the term "pioneer communities" is also

used to describe and analyse the practices of data experts in general, and Estonian data experts in particular, as influential actors in a datafied society. In this thesis, data experts are considered experts who are either responsible for data-based analyses (i.e. analysts) or build data-based solutions (i.e. developers). They are mainly considered data experts through their particular job descriptions.

Similar to the above-mentioned other influential groups, data experts share the desire to change society through data technologies All of these communities share a belief in the opportunity for productive change in culture and society through digital media and are dedicated to promoting such change (see Hepp, 2016). Therefore, all of them are by nature media-related communities, as none of them is conceivable without technical means of communication.

Although they may have a similar desire to change society, the ways in which these groups are formed vary. Data experts as a group are related very directly to their profession. These professional communities often can be more impacted by institutional barriers in realising the data-based futures they themselves as experts imagine, even if these barriers are not well-suited to the organisation. For example, many data experts have left Google because of differing views of data futures (e.g. the Timnit Gebru case).

Those communities can be considered pioneers in two ways. Hepp (2016) has argued that their self-perception is that of "pioneers," and therefore they understand themselves as being "ahead of the times." This self-awareness also fosters a sense of mission and drives pioneers to build "bridges" in an effort to bring about media-related change. Hepp (2020a) does not see those collective actors as simply "innovators" or "early adopters" (Rogers, 2003: 261–266), as they don't have the financial means to adopt technologies that ultimately fail, and those collectives also don't introduce new ideas into a system as "role models."

Pioneer communities, according to Hepp (2016; 2020a; 2020b), employ experimental practices that move beyond more established ways of doing things in their prospective domains. However, contrary to the other pioneer communities mentioned above, this research **(STUDIES III and IV)** argues that data pioneer communities cannot be considered coherent groups of actors. Data pioneer communities should always be discussed in the specific context in which they emerge, as this clearly has a major impact on how specific data technologies are imagined. As pioneer data communities emerge in private, public and other sectors, the change in stimuli in data regulation and practices certainly varies regarding their (commercial) imperatives, and this leads to power disparities among them. Market actors, like data experts from private companies, can, as DeNardis (2019) has noted, often benefit from fluid allocations of power and responsibilities, to the detriment of communities and civil society organisations.

Hepp (2020a) has also referred to pioneer communities as intermediaries in connecting developers with users, arenas of research, politics, journalism and the economy. In this, they tend to act as intermediaries between their specific domains and the public. The data experts considered in this cover article (and in

STUDIES III and IV) include a wide range, from analysts in third sector institutions to those who actually write the code for algorithms. They are all part of the process which results in specific outputs. Moreover, on many occasions, this process related to (big) data and data technologies is not completed by one specific person. This emphasises the great importance of the shared ideals of experts with different roles in this process.

Therefore, this study posits that although data pioneers in a very broad sense could be similar to the previously mentioned (see Hepp, 2020a; Hepp, 2020b) Hacker, Quantified-Self and Maker movements, considered to be pioneer communities, they are in fact much more fragmented, can be more dependent on external factors, and are affected both by underlying ideals and how data-related solutions are imagined by the general public.

The development of data technologies by data pioneers is also built on certain norms, values and ideals. Those ideals also guide the search for how to respond to the biases and unwanted consequences previous research (e.g. O'Neil, 2016; Eubanks, 2018) has highlighted. One of the responses to those worries has been to make algorithmic processes transparent and available for supervision (Diakopoulos, 2016), revealing information that is otherwise black-boxed. Others suggest thoughtful legal and technical methods for limiting those biases and consequences (Crawford and Schultz, 2014; Pasquale, 2015).

Transparency, as Ballestero (2012) has noted, is a political technology through which an attempt is made to make what would remain obscure visible. However, as discussed above, the very nature of algorithmic processes is that they remain black-boxed, and there is a point (even with particular technical knowledge) when this black box becomes impossible for any human to penetrate. Moreover, Ananny and Crawford (2018) have argued that even when we open the black boxes, this "visibility may never be sufficient to bring about a connection between author, algorithm, and consequences that is substantial enough that someone might be held accountable for an algorithm's undesirable effects."

It has also been argued (Reddy et al., 2019) that our understanding of accountability is related to the popular imagination of algorithms. This presumes that by finding out who the authors of the algorithms are, and by making them responsible parties, we could resolve the obscurity of algorithms. Reddy et al. (2019) have emphasised that "the question of accountability is caught up in normativities, shared values embedded in particular norms that legal and moral orders accept and protect." Thus, the matter cannot be reduced to authorship. Adopting an approach that builds uncritically on epistemologies that assume that "seeing a phenomenon creates opportunities and obligations to make it accountable and thus to change it" (Ananny and Crawford, 2018: 2, emphasis theirs) is not sufficient.

The ways in which these ideals are approached and realised in practice depend increasingly on different collective actors. I argue that their visions of datafied futures define the ways data is seen not just in research and organisations **(STUDIES II, III and IV)** but also on the national level.

2.3. E-Estonia digital imaginaries

Estonia, best known for its e-governance and technologically advanced digital society, where many of the services are digital, is used as an illustrative case in studying big data imaginaries by focusing on Estonian data experts' experiences and practices in building a new data-intense society **(STUDIES III and IV)** and through their practices and understandings what it means for Estonian citizens.

Since 1994, when the first Estonian IT strategy was drafted, many advances have been made. By 2021, almost all Estonian citizens had e-IDs, and approximately 70% of the citizens used them in their everyday lives (Kerikmäe & Pärn-Lee, 2021). Estonia has also contributed a lot to reducing administrative bureaucracy by using e-solutions in decision-making procedures. Today, many, if not most, public services are available online.

One of the advances which have supported and been the backbone in developing many of the solutions was the implementation of the national integration platform X-road. X-road, the core of government information systems integration, was implemented in 2000 (Dreyling et al., 2021), and since then it has played an important role in developing several public e-services, private sector services, and semi-public and semi-private systems, such as e-health systems. Many of those systems have been developed by the same private sector partners.

Most Estonians are used to technology being integrated into their everyday lives (Kalvet et al., 2013) and, compared to other European countries, are less worried about how their data are used (Murumaa-Mengel, Laas-Mikko and Pruulmann-Vengerfeldt, 2014; Special Eurobarometer, 2015). Moreover, compared to other European countries, Estonia is distinguished by not just fostering and favouring digital skills and new technology adoption, but also re-enforcing them publicly (Kõuts-Klemm et al., 2017: 282)**.** Being repeatedly named one of the digitally most advanced countries has also become Estonia's "trademark," not just in branding but also in connection with valuable knowledge through different Estonian digital society experts (see the E-Governance Academy).

Similarly to other European countries (see the Algorithm Watch 2020 Automating Society report, Estonian chapter, for an overview), the Estonian government has taken several steps to support the implementation of artificial intelligence (AI) in both public and private sector institutions in recent years. In Estonia, the term "Kratt" is used in relation to AI, and it refers to practical applications that use AI to perform specific functions (Scholl and Velsberg, 2020). In Estonia (for a longer overview, see the Estonian part of the Automating Society 2020 report), the discussions surrounding the use of automated decision-making, or AI, for the moment mainly focus on the benefits these technologies will bring. The main benefits for the public sector are considered to be increased user-centredness, improved data analysis and more efficient e-governance (see Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019: 34). Moreover, the current COVID-19 epidemic has revealed the critical importance of technology for economic and health resilience (Silaškova and Pappel, 2021),

as well as the importance of cooperation between public and private organisations in developing new technologies and in the collection of data.

Since 2017, Estonian residents have been able to check the Eesti.ee website to see which institutions have accessed their data and for what reasons by using a data tracker tool (e.g. Plantera, 2019). By September 2019, four major government agencies were participating in this project (the Population Register, the Health Insurance Fund, the Estonian Unemployment Insurance Fund, and the Social Insurance Board). The website also indicates if data have been accessed for automatic processes or by a specific automated service. The data tracker is seen as vital in fostering trust and transparency in governmental services (e.g. Plantera 2019), but by 2021 it was still not used by all of the governmental agencies. Therefore, a comprehensive overview is not available. In recent years, the Estonian State Information System Board has also been developing a Consent Service (https://github.com/e-gov/NT), which enables the data subject to give consent to a third party for the use of personal data (Kivi, 2020). Although the aim of this service is to give data subjects the opportunity to give consent before their data are shared with private entities and to foster more transparency, it also clearly highlights how governmental bodies rely increasingly on private entities to offer or develop certain services.

Kerikmäe and Pärn-Lee (2021) have emphasised that the e-governance system has been limited to Estonia, but in the case of AI solutions, global, regional and superpower interests are all important and ideally will require the same standards when applying algorithms-based decision-making and prediction models. Authors (e.g. Kerikmäe and Pärn-Lee, 2021) have also argued that because of the different levels of digitalisation in the Estonian public and private sectors, it is primarily the public sector that is encouraging others to develop and implement AI-based solutions ("use cases"; see the kratid.ee webpage for an overview of use cases). The public sector has also been advised to commission these solutions and make them publicly accessible. Thus, the Estonian AI strategy (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019) sees the public sector as a pioneer in these solutions, but imagines the private sector taking the main role in developing solutions in the future (Puusalu, 2020).

The proposals and Estonia's AI action plan (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019) focus on developing the basic competencies required for the implementation of AI. Moreover, it is emphasised in the planning document and on the e-Estonia webpage (Plantera, 2017) that "Estonia could become a role model and laboratory for the rest of the world as a place where Kratt, or AI, is put to work for the people's well-being in both the public and private sectors." Therefore, Estonia envisions data technologies as ways to further enhance digital governance.

There have also been critical voices that have stated that the projects public sector institutions take on are too narrowly focused and do not consider the wider picture needed in developing a digital society (Kerttunen, 2020). As Puusalu (2020) has noted, developments regarding AI are much needed, but the

aims proposed in the Estonian AI strategy also may lead to developing solutions which are impressive but not what are actually needed. This may lead to a situation where the development of solutions is not carefully considered and the focus is more on technologies and not on the particular needs of Estonian society.

As the previous chapter highlights, some of these developments can be considered as following the dominant imaginaries mentioned before: digital technologies and big data are imagined to help in gaining advantages and making processes more efficient. At the same time, by developing various solutions that value data subjects' role in these processes, pioneer communities developing these solutions also propose counter-imaginaries.

2.4. Research questions

To conclude, as the theoretical framework has highlighted, the emergence of big data has produced many different big data imaginaries where data is seen as a valuable resource through which more efficient and better governance can be achieved. Importantly, the ways in which big data imaginaries are realised by different actors also change data relations, where often the individual has limited agency and control over data or how they are used by institutional actors. At the same time, the new technological affordances also require individuals to take the role of data consumers in analysing and sharing their own digital data (e.g. through self-tracking solutions). This thesis aims to analyse how these imaginaries are actualised and elaborated by data pioneers and how this has affected scholarly practices. Those imaginaries and the ways they are realised affect individual and institutional agency, based on commercial and value imperatives and technological affordances. This thesis is guided by the following research questions:

1) **What are the dominant imaginaries of big data expressed by data pioneers? (STUDIES II, III and IV)**
2) **How do the dominant big data imaginaries challenge (the individual and institutional) agency (of pioneer communities)? (STUDIES I, II, III and IV)**

3. METHODOLOGY

In this chapter, I will give an overview of the data collection that was carried out for this dissertation to answer the previously stated research questions. In order to answer the study questions, I have made use of different quantitative and qualitative methods. The use of multiple methods fits with the approach of pragmatism (Creswell and Creswell, 2017: 47), which supports using different methods for the purpose of understanding a specific problem. The use of mixed methods in this study was needed to generate a better understanding of the big data phenomenon. Using multiple case studies has enabled me as a researcher to investigate, illustrate and analyse big data imaginaries and challenges to agency from both the individual and collective perspectives and in different contexts (Merriam, 2009). In the next chapter I will discuss in more depth the methods used for **STUDY I**, **STUDY II, STUDY III and STUDY IV.**

3.1. Systematic literature review method

STUDY II was based on a systematic literature review method. It mainly focused on the period of the biggest increase in the social media big data (SMBD) studies (2012–2016), in which the main shifts in the used analysis methods and knowledge-creating practices were expressed. To explain the shifts in research practices as directly reflected by the authors of these studies, a close reading of empirical social media big data studies was undertaken. The sample included peer-reviewed empirical articles which used social media big data as the main source in their research. The sample was constructed by using standardised search criteria: the concurrent search words "social media" and "big data," full-text articles accessible online, articles published only in peer-reviewed journals and written in English, and articles published in journals listed in the social science citation index. All of the articles were downloaded to a computer, and one additional search in the same database with the same keywords was done in 2017. However, some of the studies published during this period may not be represented in the sample, as some of the journals that are currently referred to as important sources for publishing peer-reviewed empirical research about (social media) big data were not indexed in the chosen database at the time of the study (e.g. the journal Big Data and Society).

The initial sample consisted of 478 peer-reviewed SMBD articles published in 2012–2016. After implementing several inclusion criteria (e.g. not consisting of SMBD analyses, essays, etc.), a final sample of 120 articles that were published between 2012–2016 was further analysed. The sample was downloaded in 2016, and all of the analyses were conducted between the years 2016 and 2018.

To study the analysis practices in the empirical articles using SMBD, a semi-structured schema was developed. This schema was tested on a smaller test sample, and then certain categories were included and clarified. I coded all of

the articles through the close reading of the full texts. The codes used for close reading reflected the identifiable formal information that was implicitly available to readers or explicitly expressed by the authors of the articles.

After closely reading and coding all of the articles, a mostly qualitative approach was used in which the thematic variations between and within particular codes, characterising single aspects of research practices, were compared systematically. Qualitative thematic analysis techniques and the software program MAXQDA were used to summarise open-ended textual codes. This also made it easy to share research notes and coded data with the co-authors of this study.

In addition, inter-coder agreement for testing the reliability of the quantitative coding was established by the co-authors rereading 11 articles from the sample. The code existence was 93.59% and the code frequency in the compared documents was 92.8%; the segment agreement was r = .82, including an average Kappa coefficient of 0.81.

The results of the main codes were additionally summarised quantitatively by using uni- and multivariate statistical techniques with an R software environment. The coded textual data were also analysed quantitatively to generalise the main differences in research practices across formal article characteristics. The differences in distribution were studied both quantitatively, using association coefficients of Cramer's V and analysis variance test F, and qualitatively, by detecting the code intersections.

3.2. Quantitative survey method

For **STUDY I,** data from the fifth round of the representative population survey "Me. The World. The Media", carried out by the Institute of Social Studies, University of Tartu, and the Saar Poll market research company at the end of 2014, were used. The survey was conducted in two languages, Estonian and Russian. The survey was carried out by the University of Tartu in cooperation with survey companies every three years (2002, 2005, 2008, 2011 and 2014) (Masso et al., 2020). This survey covered the Estonian population 15 to 79 years old, with a total sample size of 1503 (1028 respondents completed the questionnaire in Estonian and 475 in Russian). This is a thorough survey, which consists of more than 600 variables and covers the Estonian population's interests, values, traditions, religious beliefs, history, sense of belonging, identity, use of time, cultural interests, environment, consumption, health, media use, work and family. This survey made it possible to analyse how individuals (Estonian- as well Russian-speakers) cope with macro-level structural changes (Masso et al., 2020: 6) by employing single usage and a range of aggregated index variables (in 2014, there were 128 indices). The "Me. The World. The Media" survey is openly accessible through the repository datadoi.ut.ee (http://datadoi.ee/handle/33/156).

STUDY I focused on analysing a group of variables measuring perceived dangers to privacy. The following question was formulated in the questionnaire:

Have you ever had the feeling that the following institutions, companies, or persons are violating your privacy by using the Internet or social media? Next groups or institutions were presented to the respondents, who were asked to evaluate them on a five-point frequency scale, where 5 means "constantly," "very often," and 1 means "not at all"; the groups and institutions were: state institutions, local governmental institutions, employers, business enterprises, the health system, the educational system, foreigners, friends and acquaintances, and family members. As it aimed to study the relationship between privacy concerns and institutional trust, it also examined the following independent variables: *various institutions and groups essential in Estonia are listed. Please choose in each row one number from one to five that best characterises their trustworthiness for you.* Respondents were presented a list of 13 groups or institutions, including societal and media institutions and various influence groups. Respondents were asked to evaluate them on a five-point scale, with 5 being "trust completely" and 1 "do not trust at all." Next, based on these single variables, the following composite indices were calculated: trust in representatives of governmental institutions (including such institutions as parliament, the Estonian state, politicians, the president and officials), trust in other state institutions, trust in media institutions (television channels and radio stations of Estonian Public broadcasting, private television channels and radio stations, newspapers, Facebook and other social media and internet portals), and trust in cultural and surveillance institutions (schools and educational systems, the health system, the court system, banks, police, cultural activists, scientists, public companies and churches).

Besides the main dependent variable of privacy concerns, and trust in institutions and groups as the main independent variable, several socio-demographic control variables were used in the analysis: gender, ethno-linguistic affiliation, education, income and self-estimated social status. In addition, index variables characterising social media use were included in the analysis: self-expression and communication-centred internet use (uploading photos, following friends on social media etc.), use of various social media channels, being concerned about mobile or smartphone overuse in the vicinity and the functional versatility of social media use. Additionally, enterprising quality, i.e. the general activity of a person, as well as the economic activities that characterise them, was used as a background control variable.

For analysis purposes, principal-component factor analysis with Varimax rotation technique was used to reveal underlying relationship patterns among the privacy concerns regarding a list of institutions, groups and individuals. Then, individual factor scores were calculated, and mean scores across age groups were analysed to compare age groups in terms of their privacy concerns. Finally, a generalised linear regression analysis was used to explore the relationship between privacy concerns, socio-demographic variables, social media use and perceptions of social trust.

As the data for **STUDY I** were first collected during 2014, when the discussions surrounding datafication, data-driven practices and problems with algo-

rithms were not that prominent, especially in Estonia, the survey did not contain any specific questions on these topics.

3.3. Q-methodology and in-depth interviews

STUDIES III and IV were based on 24 in-depth interviews conducted as a part of Q-methodology with Estonian data experts who worked with mobility data**.** Q-methodology is a method that is often used to reveal subjective attitudes and perspectives (Stephenson, 1953). Q-methodology was used to explore the data experts' individual positions on public discourses related to data-based mobility governance. In Q-methodology, participants provide their viewpoints by ranking a set of items according to a subjective dimension, such as agree–disagree or important–unimportant, while thinking aloud (Paige and Morin, 2016). Q-methodology, as a mainly qualitative method, does not attempt, according to Van Exel and de Graaf (2005), to infer from a sample of people about an overall population, but instead selects a set of statements to represent a larger population of all possible opinions on a certain topic.

The sample consisted of 24 experts who designed data technologies for decision-making purposes in the field of human mobility. Purposeful sampling principles of the qualitative approach were followed to design the sample involving homogeneous and heterogeneous properties. The respondents worked in public, private and third sector (research) organisations. The sample was diverse in terms of gender, educational discipline and age. Three groups of data experts were represented in the sample: 1) analysts, 2) managers of analysis divisions, and 3) developers of software and algorithms. Experts' experience with mobility data varied from traditional register data to more recent digital trace data on migration. Estonia, as described in the previous chapter, offered a useful context as it shows high readiness to design and use data technologies to manage and control varying types of human mobility. As Q-methodology generally aims to establish the existence of perspectives (Brown, 1980), these studies are usually not interested in head counts or generalising to a population of people. Therefore, Q-methodological studies typically do not require large samples.

All interviews as a part of Q-methodology were carried out in Estonian during spring 2018. Except for one interview, all of them were carried out face to face. For this purpose, a semi-structured interview plan was developed. Open-ended questions, as well as projective techniques, were used to encourage experts to describe their experiences using algorithms and their understanding of algorithmic governance. Experts were asked during the interviews to reflect upon some of the most typical examples of algorithmic governance implemented in the field of migration. Such examples as a matching algorithm for refugee settlement, algorithms used in policing and the Twitter chat-bot and social media filter bubble were described. I was involved in conducting four of the 24 interviews and later was involved in the analysis of the interview data.

All interviews were recorded and transcribed. The extracts used in **STUDY III** and **STUDY IV** were translated from Estonian to English.

Additionally, as a part of the Q-methodology, all of the participants in **STUDY IV** were asked to rank and sort a series of statements (n=24) related to mobility governance through data. The statements were based on claims published in political and strategic documents, public discussions in the media and academic research. Those public discussions on data technologies covered two main conceptual dimensions: 1) subjective views on data technologies, including individual experiences or activities perceived in their organisations or society in general, and 2) the possible socio-cultural consequences of data technologies that participants considered to be negative, neutral or positive. The study introduced a conceptual matrix that was formed as a result of the intersections of these dimensions. In general, 40–80 statements are considered a standard number for a Q-sample (Watts and Stenner, 2012). In this study, equal numbers of statements were extracted in each matrix field and therefore there were a total of 36 statements (n=4 of a total Q=36).

To examine the subjective views of data experts, data were collected in three phases:

1) Open-ended questions were asked to encourage experts' spontaneous opinions on big data.
2) Data experts were asked to sort a deck of 36 cards on an agreement scale, where -5 indicated complete disagreement, +5 complete agreement, and 0 neutrality, while thinking aloud about their decision-making criteria. To encourage the expression of subjective positions, the 0-point on the scale was used conventionally, rather than "forcing" participants to choose on a given symmetrical scale.
3) Finally, data experts were asked to justify their Q-sorting choice. Usually, they were asked to justify the choices at the ends of the scale. This process lasted on average 90 minutes.

Next, all Q-Sorts were then compared and contrasted through factor analysis with a centroid rotation method to extract the discourses on data technologies. This method is preferred by most Q-methodologists as it makes it possible to explore data through rotation until the best factor solution is achieved (Watts & Stenner, 2012). The free software Ken-Q Analysis (Banasick, 2019) was used to quantitatively analyse the discourses that emerged from the data.

Then all quantitatively found discourses were compared with the use of thematic analysis techniques. The texts were grouped into meaningful categories, and text extracts within and between categories were compared. For **STUDY III,** interviewees' responses to spontaneous questions about the use of algorithms in governance, and comparisons and arguments expressed about the presented cases where the algorithmic approaches were used for managing immigration mobility were considered.

For **STUDY IV**, similar to qualitative discourse analysis, differences within and between discourses were systematically compared and interpreted, including the background information of each interviewee and the context of the

topic. Also, both manual and computer-aided techniques were combined to organise the spontaneous opinions on predefined statements, to merge the spontaneous statements into meaningful codes, and to compare the participants' positions within and between discourses. The texts were grouped into meaningful categories, and text extracts within and between categories were compared. Open-ended interview questions were analysed by using MaxQda software, and these also formed the basis of **STUDY III**.

Having discussed the methodological foundations of this thesis, in the next chapter I will present the results from **STUDIES I, II, III** and **IV,** grouped by themes.

4. FINDINGS

The empirical findings of the four studies are introduced in three parts, each consisting of subsections. I will start by describing the dominant big data imaginaries through the opportunities ascribed to big data and continue by analysing the main barriers which prevent realising these future visions. Next, I will discuss the challenges to both individual and institutional agency in a datafied society.

4.1. Main big data imaginaries

Big data have come to impact nearly all aspects of social life, from self-tracking to border management. Therefore, many of the opportunities appropriated to big data come from the widely circulated dominant big data imaginaries, which see big data as an opportunity and a valuable resource for different purposes and different actors. I will discuss these opportunities mainly based on the ways in which big data and data technologies were described by the data experts, as well as by considering the results from a systematic literature review study important in analysing how big data have changed scholarly practices.

During the last decade, scientists, and social scientists in particular, started to view SMBD as an opportunity to better understand social complexities and study human behaviour (**STUDY II).** A systematic literature review of empirical peer-reviewed articles of SMBD research indicates that although the methodological toolbox of researchers in the field of social sciences increased considerably between 2012 and 2016, their skills and suitable tools for gathering, cleaning and analysing vast amounts of social media data were still insufficient (**STUDY II)**. In analysing authors' disciplines, it became evident that this growing need for new methodologies, tools and practical skills for analysis supported the formation of different interdisciplinary teams and centres (in 17 articles), e.g. social scientists teaming up with computer scientists. Moreover, as findings from **STUDY II** revealed, interdisciplinary teams used computational techniques somewhat more often than mono-discipline teams did. In some cases, based on the researchers' comments in articles, the wish to analyse big data indicated that they were pushed into developing new tools through which social media big data could be analysed. This indicates that the availability of big data sources and interdisciplinary cooperation fostered innovation in the development of new methods and tools **(STUDY II).**

The findings of **STUDY II** also suggest that as the tools were developed and cooperation between different disciplines formed, scholars were also more eager to experiment and combine various social media datasets collected from different platforms (Twitter, Facebook, Flickr, Wikipedia etc.). The analysis of the empirical studies forming our dataset revealed that usually a range of data sources were used to compare research results or to test tools or software. Besides innovations in methods and tools, early research from SMBD studies also contributed to methodological developments and content-related innovations.

However, findings also suggest that epistemological innovations occurred after the new analysis techniques and methods had been developed and implemented **(STUDY II)**.

Besides by scientists, datafication has been seen as an opportunity by Estonian data experts, as reported in **STUDIES III and IV**. The interviewed data experts **(STUDIES III and IV)** described datafication as an opportunity to enhance governance by improving decision-making and increasing efficiency. In data experts' opinions, increasingly data as resources were considered helpful in achieving competitive advantages for both private and public organisations. Results from **STUDIES III and IV** indicate that, in some cases, data were also treated as a commodity that could be sold. This highlights and supports the dominant big data imaginaries in which data are often considered a capital.

The interviewed experts also expressed the need for global uniform standards, as without them, in their opinions, the full potential of digital data could not be achieved. Thus, big data are not seen as merely resources to develop fair solutions within national boundaries but also as global opportunities to meet universal obligations and rights **(STUDY IV)**. Therefore, it is believed that in the future (big) data and data technologies will be equally shared between different data consumers. Moreover, as this approach towards equal standards was expressed especially in relation to mobility data, it was believed that data and data technologies would form the solutions to society's most pressing problems.

Some data experts (**STUDY IV**) considered data to be a human right, which means that both experts and members of society should have equal access to data. Data experts (**STUDY IV**) stated that the openness of data was an important prerequisite for realising many of the future pathways described by them. Open data were seen as providing an opportunity to move towards a more transparent state and governance **(STUDY IV)**. The availability of data as a human right in this context was more a metaphorical concept, which reflected specific barriers the experts had experienced in their work.

Additionally, in some interviews the use of data was considered a possible way to control certain processes. Securitisation, a rather important topic in previous research, was not mentioned frequently in **STUDIES III and IV**. Often the understanding of data as an open resource for everybody was also related to the related values of transparency and accountability. Findings from **STUDIES III and IV** related transparency mainly to the opportunity to challenge decisions based on data. Experts related data to accountability among decision-makers and through them institutions and not so much to individuals' digital literacy. The previously mentioned data tracker used in Estonia supports this view of transparency shared by experts, as it allows individuals to see which of their data have been used by which public institutions or private organisations. Access to data, as mentioned above, is seen as an opportunity to also have a more transparent state and governance. Part of this transparency, as data experts emphasised, involves system audits that could help prevent possible problems related to data bias **(STUDY IV)**.

Thus, the realisation of these dominant big data imaginaries, which see big data as an opportunity to better deal with social problems and to develop more efficient governance, is blocked by different barriers.

4.2. Barriers preventing imagined data futures

The results from **STUDIES II, III and IV** also revealed several barriers and risks associated with increasing datafication. One of the barriers both addressed by data experts **(STUDIES III and IV)** and described in peer-reviewed empirical articles **(STUDY II)** was access to data. Barriers to accessing data are connected with several different factors, including the lack of specific skills, technological affordances and unified standards, legal restrictions and data relations.

The lack of specific skills is an important factor for both SMBD researchers **(STUDY II)** and data experts **(STUDIES III and IV)**. The lack of certain skills, together with the lack of suitable tools to analyse SMBD (**STUDY II)** can be associated with the emergence of digital methods initiatives that aim to develop methods and tools for big social data analyses and winter and summer schools (e.g. the Digital Methods Initiative in Amsterdam) that foster cooperation between different researchers. Although the lack of skills is a barrier in using SMBD, it has fostered both disciplinary and interdisciplinary cooperation and the development of new tools and methods described in the previous chapter.

While the development of skills is increasingly supported through the emergence of several initiatives and cooperation, specific technological affordances of the platforms have remained problems. More than half of the studies in 2012–2016 that were part of the sample in **STUDY II** used Twitter data. Compared to other social media platforms, Twitter has remained more accessible than Facebook and has allowed researchers to access its data through its APIs (Application Programming Interfaces). Additionally, there are easy to use end-user tools to capture data from Twitter and analyse them. These tools have certain limitations, such as the volume of data that can be analysed and the number of features available for the analysis. This also illustrates the changing dynamics in data relations, where researchers are not dependent on respondents to answer specific survey questions, but may not have access to data because of platform owners' policies and technological affordances.

While researchers have been mostly dependent on specific technological affordances provided by platforms, findings from **STUDIES III and IV** suggest that traded data relations, in which an organisation is financially dependent on customers, form another obstacle to novel data solutions **(STUDY IV)**. As findings from **STUDY IV** indicate, dependence on customers' expectations is especially important for third-sector organisations. This means that data experts are reliant on the ways that data and data-based approaches are imagined by their clients, not how they themselves imagine them. In some cases, even if

experts expressed ways in which data-based solutions could be used to solve complex problems, such infrastructural and legal restrictions as access to data and differences in standards formed barriers. In many instances, the interviewed public sector experts had a good general understanding of what kind of data other public sector institutions might have and which could be useful for their analyses. Problems, however, arose when there were differences in standards or specific data could not be shared because of legal restrictions connected with data protection rules (**STUDY IV**). Interviews with data experts **(STUDY III)** showed that even governmental institutions sometimes refused to share data because of possible advantages for their own organisations that the data could hide. Some experts in the sample had also tried to cooperate with private companies, but in most cases those attempts were not very successful **(STUDIES III and IV)**. The interviewed experts believed that private companies not sharing data with public institutions was usually justified through concerns for privacy **(STUDY III).** Accordingly**,** the findings of **STUDY I** revealed that the privacy concerns of Estonians are related to institutional trust and, therefore, institutional practices regarding data and data technologies have a clear impact on how trustworthy those organisations are perceived to be by the public.

4.3. Challenges for agency and changed data relations

The developments related to the new data technologies have also challenged the agency of both individual and institutional actors and changed data relations. Based on the results of my studies (**STUDIES II, III and IV**), I will approach agency through the three types of roles different individual and institutional actors play in the data ecosystem: data producers, data distributors and data consumers. These terms are used here to describe the data relations between these actors and the main challenges to their agency.

4.3.1. Challenges for data producers

Data producers fuel the process of datafication through their data. Data producers here include both individual and collective actors. Through digital footprints left by individuals, with or without acknowledging them, through using different digital services, they insert data which are later analysed and used by data consumers. Through their own self-tracking practices, users may take the position of data consumers.

However, according to the data experts (STUDY IV), as data producers individual users lack control over what happens with their data. Several data experts argued that, at the individual level, it is hard or even impossible to control what happens with one's data, as many of the processes and uses may remain unknown to the user. This was especially highlighted in the case of private companies, whose commercial imperatives are also fuelled by the secrecy of their services. Similarly, users may not know if researchers have

gathered their tweets, posts or Instagram pictures to study certain topics. Today several organisations (e.g. AoIR) have formed ethical guidelines for research that analyses data from social media platforms. However, these guidelines had not been fully established in the beginning years of SMBD research.

Whatever the specific aims which organisations use to justify data collection and the use of data technologies, as control over one's data has become much harder to obtain, the current measures (e.g. notice and consent) for privacy protection do not result in individuals controlling their privacy. This is supported by the results from **STUDY I,** which showed that Estonian individual privacy concerns are related to institutional trust, as emphasised in the previous chapter. In a datafied society where the use of data technologies has made many processes opaque to individuals, the main privacy concerns are related to institutions, so the importance of trust individuals have in certain services has increased.

This institutional trust is achieved and also challenged through specific technological affordances. The notice and consent option is related to interfaces and options provided by organisations. Therefore, the old model tends to be integrated into the new data-centric ecosystem, where it does not provide options for individual agency.

Previous research has shown that big tech companies, even when using data for the social good, are mainly guided by their commercial imperatives (see Magalhaes and Couldry, 2021), where in order to gain access to certain services, platforms etc., individuals have no other choice than to share their data. Public sector institutions' initiative, based on **STUDY III,** differs from that of commercial entities in that, rather than producing value through data for their organisations, other parties or individuals, they control which data are collected from individuals and for which purposes. The Estonian public sector is built on digital infrastructure and data. Those data-centric practices help to provide individual citizens with proactive services. In those services, citizens' data is used to analyse whether they have rights to certain services and the services are offered without an individual applying for them. However, this practice includes the *systematic coercion of digital participation* (see Barassi, 2019), as to obtain certain services individuals need to provide data to public entities. Therefore, although in the public sector there may sometimes seem to be more freedom of choice for the individual, in societies that are more and more based on those data-driven analyses and services the choice is rather illusory.

However, individuals can also be empowered through certain measures which acknowledge their role as data producers in a datafied society. Estonia uses a data tracker **(**Plantera, 2019**)** through which citizens can see which organisations have used their data, and they can request more information about those uses when needed. This not only increases institutional trust by reducing privacy concerns but also gives individuals the opportunity to challenge institutional practices and emphasises that data producers are also data owners.

4.3.2. Challenges for data distributors and data consumers

Standing between data producers and data consumers there are usually data distributors. I define data distributors as actors who maintain certain services through which different parties can input data and collect it for further analysis, but they are also often the ones who develop data technologies and, in this role, choose technological affordances for both data producers and data consumers. Estonian data experts **(STUDIES III and IV)** can be considered both data distributors and data consumers, depending on their organisations, their roles and the work they do in their organisations. SMBD researchers are, in this cover article, mainly considered as data consumers **(STUDY II),** as the studies analysed through the systematic literature review focused on their practices related to gathering and analysing social media big data**.**

As the previous chapter argued, private companies are today mostly guided by the commercial imperative: the collection and analysis of data are considered actions through which their financial profits can be increased. The findings suggest that data experts **(STUDIES III and IV)** and researchers who make use of social media big data **(STUDY II)** are guided by the value imperative**.** In several interviews with data experts **(STUDIES III and IV)**, it was emphasised that the main aim of their work was, by using data, to offer value to their clients and/or society. Data experts are active in both explaining and actively offering data-based approaches to analysis for their customers or partners (**STUDY IV**). Some experts realise that by translating the needs of the customer, they "value data." Value creation refers here to the fact that experts see data as important input for better decision-making on many levels. Thus, not all experts see having vast amounts of data or even accessing data as the most important way this value can be achieved. Some of the experts expressed the importance of more qualitative methods, such as interviews, to collect better data through which individuals' needs could be served. This indicates that experts play an important role in balancing the needs of their customers, who can be either other departments in their own organisation or other organisations from the private or public sector, with the needs of data subjects, e.g. data producers in the context of this cover article.

Findings from **STUDY IV** also highlight how, in some cases, data experts felt the need to take a very active role in introducing data-driven approaches, e.g. a new kind of culture, data culture, within their organisation **(STUDY IV)**. Some of the data experts who actively participated in introducing this new type of culture referred to themselves as data enthusiasts. These experts are especially driven by imaginaries in which data are considered the key components in achieving efficiency, better control and the social good. Moreover, those experts said that through developing this culture in their organisations, they also improved the state (**STUDY IV**). In their opinion, data technologies in governance offered opportunities to support the progress and reputation of the state. Therefore, by actively introducing novel data analysis tools, they had promoted and contributed to a shift towards a data-oriented culture. However, being a data

philanthropist who develops the data culture comes, as the findings from **STUDY IV** indicate, with a responsibility to the public. Therefore, some data experts also referred to the necessity to introduce regional or global governance principles to protect human rights in this novel situation where data instead of humans are on the move across national territories **(STUDY IV).** Therefore, experts are not just data consumers who analyse data and develop new data solutions, but they also play a very active role in much bigger transformations in societies and, therefore, can be considered data pioneers who form data pioneer communities**.**

Similar to the results regarding data experts **(STUDIES III and IV),** the results from **STUDY II** indicate that SMBD is considered a valuable resource for studying aspects of human behaviour and actions through data collected by social media companies. Value is, however, not just achieved through the analyses and new knowledge produced from these analyses, but also, increasingly, through the methodological and other innovations described in the previous chapter.

However, researchers are constrained by the interface through which companies provide access to SMBD **(STUDY II).** Researchers have to be constantly ready for changes in those interfaces, which are mostly guided by the commercial or other imperatives companies have. For example, Twitter, which was mostly used according to the findings of **STUDY II**, has changed its API many times and, with it, its rules. Also, through technological affordances, Twitter controls which data can be accessed: the length of the specific period and which data researchers have access. This also emphasises the power relations which control today's datafied society, where monopolised Big Tech has power over most of the data through which the social good can be developed but which can only be accessed by a few. Findings from **STUDY III** also reveal that the lack of access to data and lack of certain skills also pose challenges for data experts and may lead to inequalities in pioneer communities. Although cooperation is valued and seen as a way to enhance the social good through data, it is hard to achieve between private and public entities. Private companies, in the example given in **STUDY III,** tend to back away from cooperation and justify this through the need for privacy protection and the GDPR regulation. The GDPR regulation was implemented to give back some control to the individual, but in reality, as indicated in **STUDY III**, it also makes the processes in the public sector more complicated and leads to manual data collection, duplicate databases or inefficient actions.

The pioneer communities data experts form, however, are not coherent groups of actors. The findings from **STUDY IV** suggest that the data futures they imagine are not only impacted by regional settings and organisational culture but also by their own experiences as individual data producers. Their experiences as users of services or data subjects shape the ways they approach certain societal problems in their everyday work and which kinds of solutions they imagine with (or without) data technologies.

5. DISCUSSION

The aim of my thesis was to examine how dominant big data imaginaries are actualised and elaborated by data pioneers, and how this has affected scholarly practices and challenged individual and collective agency. For this purpose, I have used the concept of big data imaginaries, which I have defined in this thesis as the outcomes of complex individual and collective sense-making activities regarding shared ideas about data technology, which are often impacted by individual or collective ideals, ideas about technology and fears about negative societal effects.

There are always multiple imaginaries in circulation and they are more or less powerful. Some imaginaries can run peacefully in parallel, while others may conflict with each other and seek dominance or resistance (Jasanoff, 2015: 329). Although in this cover article I have not focused on the ways that the socio-cultural context of previous technological infrastructure affect the development of big data imaginaries, I consider these aspects important to consider when discussing the main results of my work.

The circulation of big data imaginaries is mostly motivated and propagated by commercial actors' assumptions about data and the ways they have chosen to design their products. Private companies' data-driven practices are mainly determined by the commercial imperative (see Magalhaes and Couldry, 2021), which treats the personal data that users produce through their everyday actions digitally as commodities and resources for economic gain. This is increasingly supported by other, often partly hidden, collective actors interacting with the data subject, such as advertisers and investors, who also aim to produce specific value through data-driven processes. The concept of making data valuable for someone or something was also central in the discussions with Estonian data experts **(STUDIES III and IV)**. Data experts' value imperative partly differs from the commercial imperative in the desire to balance institutional aims with data subjects' needs. Although these transformations are seen happening through data-driven approaches, data experts also deflect responsibility to the public while introducing these value-producing data-driven approaches. As many obstacles described by the data experts are related to data access or specific skills needed for data analysis, the overall approach to data technologies is similar to that in Broomfield and Reutter's (2021) research on Norway's public sector experts, which is also technology-oriented. This technology-centredness has been especially visible during the COVID-19 crisis, with solutions being sought through data technologies or data analysis and, as a result, there has been more emphasis on numbers than on the contexts they appear in. Although these transformations are seen happening through data-driven approaches, data experts also shift responsibility to the public while introducing these value-producing data-driven approaches **(STUDY III)**.

However, as data experts also often consider themselves responsible for developing data cultures, i.e. promoting data-driven approaches, in their organi-

sations, I consider them to also be, similar to the Hacker Movement and Quantified Self Movement (Hepp, 2016, 2020a, 2020b), a pioneer community. Although as a data pioneer community they are not a coherent group of actors, they are impacted by different imperatives and their role as intermediaries is complicated by access problems and certain power relations, they are clearly active agents whose visions have the power to steer the process along specific pathways. It is also important to acknowledge the context they emerge in as it greatly impacts the ways their ideas materialise. In Estonia, new technological developments are not just warmly welcomed and seen as the building blocks of better governance, but are also often considered important for positive branding of the state, its progress and its reputation.

This, however, also further impacts the agency of both individual and collective actors. In this cover article, I have adopted the Klinger and Svensson (2018) model of agency to discuss the three main roles agents play in a datafied society. Based on the findings of **STUDIES I, II, III and IV,** I propose a model where both collective and individual actors can act as data producers, data distributors and data consumers who are affected by individual and shared ideals, different imperatives often depending on the institutional, as well as national and international, contexts and technological affordances posed by themselves or by other actors (see Figure 2). Based on the results of **STUDY IV**, these roles are dependent on each other, and this also impacts their data-related practices and experiences in a datafied society.

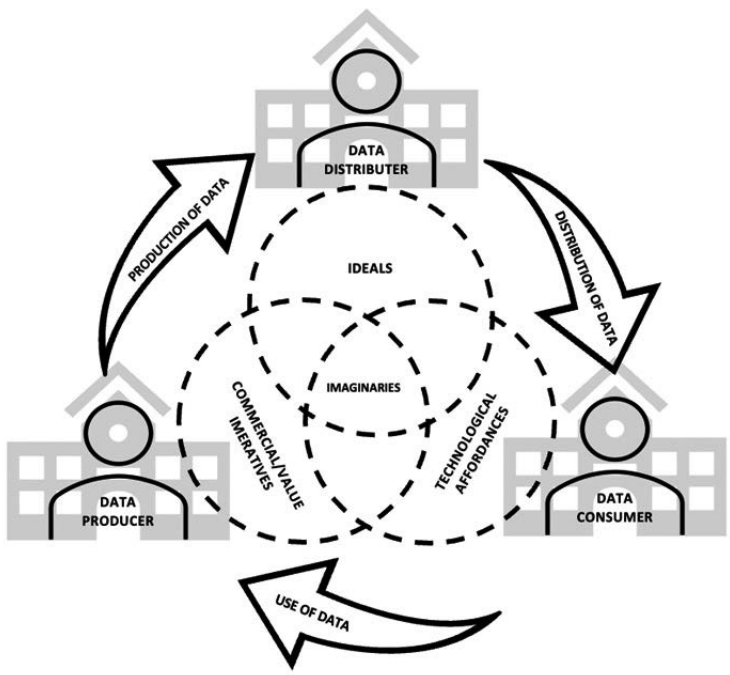

Figure 2. (Imagined) data agency challenged and shaped by the triple roles of individual and collective actors. Author's model.

In this thesis I consider these roles to be played by both individual and institutional actors. Often individual users act as data producers through their practices, in which digital traces are produced both intentionally and unintentionally. Similarly, individual users can also be increasingly seen as data consumers as they, for example through different self-tracking solutions, make databased decisions based on their own data.

Data is mainly distributed today through digital channels. Technology companies have increasingly played the role of both social and economic intermediaries by often providing essential services, products and infrastructures in exchange for access to people's personal data (Prainsack, 2019). Similarly, governments and public sector institutions may act as data distributors. As data distributors, these institutions are often in power positions as they can choose which data, through which channels and with whom the data are shared, and whether this sharing supports their own imperatives. Therefore, I consider technological affordances to be closely related and also often justified through the different imperatives of specific organisations. Private companies are mostly guided by commercial imperatives, where data form the basis of innovations and are needed for the companies to be competitive and considered valuable by other stakeholders. Seeing data as a competitive advantage can, therefore, help to justify the opacity of data-related processes.

Powerful organisations as data distributors are also often responsible for the technical infrastructure and the technical affordances related to specific digital environments. These specific technological affordances, as **STUDY II** has shown, affect knowledge production in the (social) sciences. In research communities, the digital traces left by individuals while using (social media) platforms have been seen as ways to better understand social complexities and to study human behaviour. However, as the studies published in 2012–2016 demonstrated, the platforms themselves somewhat restrict access to data, or make it complicated to collect data without specific tools and technical knowledge (**STUDY II**). Therefore technological affordances posed by data distributors often increasingly push other actors, e.g. data consumers, to learn new skills or develop software and tools, as well as to produce new kinds of digital divides (Andrejevic, 2014). This situation illustrates the change in the dynamics in data relations as researchers **(STUDY II**) are increasingly dependent on platform owners' policies, technological affordances and their own capabilities, rather than data subjects' willingness to participate. Often data subjects have no clue that they have been part of studies, and that their likes or public messages on different feeds have been interpreted in certain ways.

Individuals, as data producers, are also often systematically coerced to digitally participate as the choice not to share specific data with (public sector) institutions results in them being removed from specific services. These processes make data producers increasingly dependent on processes, which in many cases, as research (Gillespie, 2014; Pasquale, 2015) shows, remain opaque to them. Although Barassi (2019) mainly uses this concept in relation to the public sector, data subjects are also increasingly pushed to use different

private sector services. Especially in 2020 and 2021, during the COVID-19 pandemic, many people have had to work from their homes; people have also been coerced to use many private sector solutions for their work: Zoom for lectures and meetings, Teams to transfer specific work data, etc. As data producers, individuals often also have certain specific aims in using different services. Thus, they are not just passive agents, and their practices are often guided by their own personal imperatives and needs.

As the control over one's data has become much harder to obtain **(STUDIES III and IV)**, the currently used measures (e.g. notice and consent) for privacy protection do not result in individuals controlling their privacy. The notice and consent option, widely criticised by researchers (Bannermann, 2019), is related to the interfaces and options provided by the organisations themselves. Therefore, the old model tends to be integrated into the new data-centric ecosystem, where it does not help individuals to protect themselves against faulty and problematic institutional practices. This model, however, offers institutions a way to continue to concentrate on questions of data collection and not so much on the ways data are used in these systems. Data collection may be regulated through legislation and, through this, individuals can be protected from some of the consequences, but the ways data are used may remain vague. Moreover, as Broomfield and Reutter (2021) argue, privacy and legal issues are often linked with trust, which means that at least on the public sector level if privacy is protected and current regulations are adhered to there is little concern.

Certain technological affordances can also act increasingly as possible ways through which institutional trust (**STUDY I**) and transparency (**STUDY III**) as ideals are fostered. For example, in Estonia, a solution called Data Tracker (see Plantera, 2019) aims to make the institutions that have analysed citizens' data more transparent to citizens. Although it aims to increase the institutional trust valued by data subjects (**STUDY I**), it also produces an "illusion of transparency", especially in the public sector context. These kinds of technological solutions are imperfect, which is apparent when someone tries to test the decisions which have resulted from these automated analyses. This part of the process, however, is in many cases made increasingly uncomfortable for the data subject. But users' power to control or test these processes is very much limited in datafied societies **(STUDY III)**. Therefore, as control and power are in the hands of collective actors, such as data distributors and data consumers, this is not about the trust individuals have in certain entities but rather about the trustworthiness of the data-related practices of these actors.

This study has its limitations. The limitations of this thesis, however, offer promising and interesting opportunities for future research. In discussing the possible ways that datafication has affected scholarly practices, I have relied on peer-reviewed articles and researchers' reflections on articles. However, their reflections are often limited based on the specific journals these studies were published in. In future studies, interviews or other qualitative approaches could offer better insights into people's experiences with big data, as well as the challenges they encounter.

This thesis has mainly analysed big data imaginaries, as well as the challenges the data ecosystem poses to data experts. It considers that certain ideals data experts share as pioneer communities have important impacts on how data culture is fostered in organisations, as well as on how certain solutions are approached. In approaching data experts as pioneer communities, they could also be considered to operate as interconnected infrastructures (Van der Vlist and Helmond, 2021), where different actors have different infrastructural power (Van Dijck et al., 2019). However, this ignores other influential actors whose practices may impact and foster future imaginaries. Therefore, I suggest that future research could also focus on mapping which kinds of actors, often somewhat ignored in discussions, have bargaining power when developing data-driven systems. Moreover, I suggest that including investors, an important and often ignored group by data producers, could help clarify why certain processes are constantly repeated and why imagined changes are so hard to achieve.

6. CONCLUSIONS

In this section, I provide answers to the research questions and highlight the empirical contribution of my thesis.

1) **What are the dominant imaginaries of big data expressed by data pioneers? (STUDIES II, III and IV)**
   a) Big Data is mainly imagined as a valuable resource that provides means to enhance governance through improved decision-making and increasing efficiency. Data are increasingly treated as a commodity and capital considered to provide an important competitive advantage for both private and public sector organisations. This clear advantage may impede some collaborations as organisations hope to gain positive advantages by analysing certain data **(STUDIES III and IV)**.
   b) Data, especially big social media data, are also imagined to offer an opportunity to better understand social complexities and study human behaviour. In the first years of gathering and using big social media data for research, the available methodological toolbox and skills (especially in social sciences) were not sufficient, and this led to the formation of different interdisciplinary teams and centres, e.g. social scientists teaming up with computer scientists. Teaming up with scientists from fields where computational methods were used often provided better opportunities to collect vast amounts of data. Besides cooperation between different disciplines, the availability of big data sources also fostered innovation in the development of new methods and tools needed for data gathering and analyses. At first, therefore, researchers were also eager to experiment and combine various social media datasets collected from different platforms for the purpose of testing tools/software developed for these analyses. Although early research from social media big data contributed to methodological developments and content-related innovations, the findings also suggest that epistemological innovations will occur after the new analysis techniques and methods have been developed and implemented **(STUDY II)**.
   c) The opportunities imagined by data experts and in social media big data empirical research articles are often complicated by encountering different barriers. Problems with accessing data were often mentioned as the main barrier in implementing data-driven approaches. The problems described include insufficient skills and knowledge to gather or analyse (big) data, the lack of unified standards needed for sharing data between different parties, legal restrictions usually implemented to protect data subjects' rights, technological affordances and changed data relations **(STUDIES II, III and IV)**.
   d) Besides the lack of skills, which has been addressed through better cooperation between different disciplines and digital methods initiatives, the technological affordances of (social media) platforms have posed

challenges to access to data. As access to data is mostly granted to stakeholders who use data for their commercial needs (i.e. advertisers), this also highlights the changing dynamics in data relations. Data form a resource collected by institutions and shared/sold between institutions. Data subjects are increasingly treated as commodities **(STUDY II)**.

e) Changes in data relations also impacted the data experts, especially in third sector organisations, where data were traded. The dependence on customers' expectations of these organisations leads to data-based approaches, which are imagined by their clients, not by the experts. Although in some cases experts may actively try to introduce and offer new approaches to their clients, different infrastructural and legal restrictions may arise **(STUDIES III and IV)**.

2) **How does datafication challenge (the individual and institutional) agency (of pioneer communities)?**

a) The barriers that arise for data experts and researchers have inevitably challenged agency at both the individual and collective levels. Agency is challenged through different imperatives organisations have when developing data technologies, as well as by distributing this data through infrastructural components. Private companies are usually motivated by the commercial imperative, which is driven by their own aims and the aims of different, often ignored stakeholders. The findings suggest that data experts and social media big data researchers are guided by the value imperative as the aim is to provide value (new knowledge) through data analysis to customers and society. However, with both groups, it is increasingly important to balance the needs of customers with the needs of data subjects **(STUDIES II, III and IV)**.

b) Data experts imagine data as the main source for efficiency, better control and the social good, and see themselves as having a very active role in doing this. For this purpose, some of them actively introduce a new, data-driven culture in their organisations and see themselves as data enthusiasts. Therefore, experts can be considered data pioneer communities that mediate the data-based knowledge both inside their organisations and with their customers **(STUDY III)**.

c) Individuals, as data producers, based on data experts' understanding, lack control in datafied societies and, although digital literacy is important, the opacity of certain processes makes it increasingly difficult for them to challenge decisions made by institutions (**STUDIES III and IV**).

d) In introducing these new approaches, the data experts also emphasised their responsibility to the public and the importance of the related values of transparency and accountability. Transparency and accountability are mainly related to the possibility of challenging decisions based on data. The data experts expressed the importance of the accountability of organisations as individuals often do not clearly understand by whom and for which purposes their data are used. Therefore, the currently used measures (e.g. notice and consent) often do not result in individuals

controlling their privacy. As a result, the trust that individuals have in certain services, i.e. institutional trust, has become even more important (**STUDIES I and III**).

REFERENCES

Algorithm Watch, & Bertelsmann Stiftung. (2019). Automating Society: Taking Stock of Automated Decision-Making in the EU (No. 1). Retrieved from https://algorithmwatch.org/en/automating-society/

Algorithm Watch, & Bertelsmann Stiftung. (2020). Automating Society: Taking Stock of Automated Decision-Making in the EU (No. 2). Retrieved from https://algorithmwatch.org/en/automating-society/

Alston, P. (2019). Report of the Special rapporteur on extreme poverty and human rights. A/74/ 48037. Retrieved from https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25156

Ananny, M. (2011). The curious connection between apps for gay men and xex offenders. The Atlantic. Retrieved from: http://www.theatlantic.com/technology/archive/2011/04/the-curious-connection-between-apps-for-gay-men-and-sex-offenders/237340/

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20(3)*, 973–989. https://doi.org/10.1177/1461444816676645

Andrejevic, M. (2014). Big data, Big questions: The big data divide. *International Journal of Communication*, *8*, 1673–1689

Angwin, D. N., Mellahi, K., Gomes, E., & Peter, E. (2016). How communication approaches impact mergers and acquisitions outcomes. *The International Journal of Human Resource Management*, *27*(20), 2370–2397.

Ballestero, A. (2012). Transparency in triads. PoLAR: *Political and Legal Anthropology Review, 35(2)*, 160–166.

Banasick, S. (2019). Ken-Q Analysis (Version 1.0.6) [Computer software]. Retrieved from https://shawnbanasick.github.io/ken-q-analysis/.

Bannerman, S. (2019). Relational privacy and the networked governance of the self. *Information, Communication & Society*, *22*(14), 2187–2202

Barassi, V. (2019). Datafied citizens in the age of coerced digital participation. *Sociological Research Online*, *24*(3), 414–429.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, *104*, 671.

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society, 11*, 985–1002

Beer, D. (2016). *Metric power*. London: Palgrave Macmillan.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the New Jim Code. Polity Press: Medford.

Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns, 2(2)*, 100205.

Bodó, B. (2019). Selling news to audiences–a qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media. *Digital Journalism*, *7*(8), 1054–1075.

Bolin, G., & Andersson Schwarz, J. (2015). Heuristics of the algorithm: Big Data, user interpretation, and institutional translation. *Big Data & Society*, *2*(2), 2053951715608406.

Bonde Thylstrup, N., Flyverbom, M., & Helles, R. (2019). Datafied knowledge production: Introduction to the special theme. *Big Data & Society*, *6*(2), 2053951719875985.

Bowker, G. C. (2014). Big data, big questions| the theory/data thing. *International Journal of Communication*, *8* (5).
Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15(5)*, 662–679.
Braun, J. (2013). *Transparent intermediaries: Building the infrastructures of connected viewing*. In: Holt, J. and Sanson, K. (eds) Connected Viewing. London: Routledge, 124–143.
Brauneis, R., & E. P. Goodman (2018) Algorithmic transparency for the smart city, *Yale Journal of Law and Technology*, *20*, 103–176
Breiter, A., & Hepp, A. (2018). The complexity of datafication: putting digital traces in context. In *Communicative figurations* (pp. 387–405). Palgrave Macmillan, Cham.
Broomfield, H., & Reutter, L. M. (2021). Towards a Data-Driven Public Administration: An Empirical Analysis of Nascent Phase Implementation. *Scandinavian Journal of Public Administration*, *25*(2), 73–97
Broughton Micova, S. & Jacques, S. (2020). Platform power in the video advertising ecosystem. *Internet Policy Review 9(4)*, 1–28
Brown, S. R. (1980). Political subjectivity: Applications of Q methodology in political science. New Haven: Yale University Press.
Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.
Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society, 3(1)*, 1–12
Carmi, E. (2020). Media Distortions: Understanding the Power Behind Spam, Noise and Other Deviant Media. Peter Lang.
Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International & Comparative Law Quarterly*, *69(4)*, 819–844.
Cohn, J. (2019). The burden of choice: Recommendations, subversion, and algorithmic culture. Rutgers University Press.
Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
Couldry, N. (2014) Inaugural: A necessary disenchantment: Myth, agency, and injustice in a digital world. *The Sociological Review 62(4)*, 880–897
Couldry, N., & Hepp, A. (2017). *The mediated construction of reality*. John Wiley & Sons
Couldry, N., & Powell, A. (2014). Big data from the bottom up. *Big Data & Society*, *1*(2), 2053951714539277.
Couldry, N., & Yu, J. (2018). Deconstructing datafication's brave new world. *New media & society*, *20*(12), 4473–4491.
Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.*, *55*, 93.
Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.
DeNardis, L. (2019). The Social-Media Challenge to Internet Governance. In *Society and the Internet* (pp. 389–402). Oxford University Press.
Dencik, L. (2020). Mobilizing Media Studies in an Age of Datafication. *Television & New Media*, *21*(6), 568–573.
Dencik, L., Hintz, A., & Carey, Z. (2018). Prediction, pre-emption, and limits to dissent: Social media and big data uses for policing protests in the United Kingdom. *New Media & Society*, *20*(4), 1433–1450.

Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the ACM, 59(2)*, 56–62

Difi (2018) Deling av data: Konseptvalgutredning. Retrieved fromhttps://www.difi.no/sites/difino/files/deling_av_data_kvu_sladdet.pdf.

Dreyling, R., Jackson, E., Tammet, T., Labanava, A., & Pappel, I. (2021). Social, legal, and technical considerations for machine learning and artificial intelligence systems in government. In *International Conference on Enterprise Information Systems ICEIS*

Emirbayer, M., & Mische, A. (1998). What is agency? *American journal of sociology*, *103*(4), 962–1023.

Eslami, M, Rickman, A, Vaccaro, K, Aleyasen, A, Vuong, A, Karahalios, K, Hamilton, K & Sandvig, C. (2015). 'I always assumed that I wasn't really that close to [her]: reasoning about invisible algorithms in news feeds,' CHI'15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems, April 2015. https://doi.org/10.1145/2702123.2702556

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

European Commission (2015). Special eurobarometer 431: Data protection. Retrieved from: http:// ec.europa.eu/public_opinion/index_en.htm

European Commission. (2016). Data Protection in the EU. Retrived from https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Felt, U., & Öchsner, S. (2019). Reordering the "world of things": the sociotechnical imaginary of RFID tagging and new geographies of responsibility. *Science and engineering ethics*, *25*(5), 1425–1446.

Franzke, A.S., Muis, I. & Schäfer, M.T. Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands. Ethics Inf Technol (2021). https://doi.org/10.1007/s10676-020-09577-5

Gawer, A. & Cusumano, M.A. (2014) Industry platforms and ecosystem innovation. *Journal of Product Innovation Management, 31(3)*, 417–433

Gillespie, T. (2014). The relevance of algorithms. In: Gillespie, T., Boczkowski, P. and Foot, K. (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, 167–194.

Gitelman, L., and Jackson, V. (2013). Introduction, In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, 1–14

GitHub – e-gov/NT: Nõusolekuteenus / Consent Service. (2021). GitHub. Retrieved from https://github.com/e-gov/NT

Goffey, A. (2008). Algorithm. In M. Fuller (Ed.), *Software studies – A lexicon* (pp. 15–20). Cambridge: MIT Press

Green, L. (2001). Technoculture. Crows Nest: Allen and Unwin

Hammersley, B. (27.03.2017). Concerned about Brexit? Why not become an e-resident of Estonia? Wired. Retrieved from http://www.wired.co.uk/article/estonia-e-resident

Hartley, J. M., & Schwartz, S. A. (2020). Trust, disconnection, minimizing risk, and apathy. *MedieKultur: Journal of media and communication research*, *36*(69), 011–028.

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G. and Tao, R. (2015), "Gaining competitive intelligence from social media data: evidence from two largest retail chains in the world," *Industrial Management & Data Systems*, Vol. 115 No. 9, pp. 1622–1636. DOI: 10.1108/IMDS-03-2015-0098

Helbing, D. (2015). *Thinking ahead-essays on big data, digital revolution, and participatory market society* (Vol. 10). Cham: Springer

Hepp, A. (2016). Pioneer communities: collective actors in deep mediatisation. *Media, Culture and Society, 38(6)*, 918–933. https://doi. org/10.1177/0163443716664484

Hepp, A. (2020a) Deep mediatization. Oxon, New York: Routledge

Hepp, A. (2020b). The fragility of curating a pioneer community: Deep mediatization and the spread of the quantified self and maker movements. *International Journal of Cultural Studies*, *23*(6), 932–950.

Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 48–62.

Hintz A., Dencik L., Wahl-Jorgensen K. (2019). Digital citizenship in a datafied society. Polity Press: Medford

Hockenhull, M., & Cohn, M. L. (2021). Hot air and corporate sociotechnical imaginaries: Performing and translating digital futures in the Danish tech scene. *New Media & Society*, *23*(2), 302–321.

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*. 35, 441–456.

Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology, & Human Values*, *41*(1), 17–49.

Jarvenpaa, S. L., & Markus, M. L. (2020). Data sourcing and data partnerships: Opportunities for IS sourcing research. In *Information Systems Outsourcing*. Springer, Cham. 61–79

Jasanoff, S., (2015) Future imperfect: Science, technology, and the imaginations of modernity. In: Jasanoff, S, and Kim, S. (eds) *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power.* Chicago: University of Chicago Press, 1–33.

Jasanoff, S., & Kim, S. H. (2009). Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva*, *47*(2), 119.

Jasanoff, S., & Kim, S. H. (2013). Sociotechnical imaginaries and national energy policies. *Science as culture*, *22*(2), 189–196.

Jørgensen, R. F. (2021). Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society*, 1–16

Kasapoglu, T., Masso, A., & Calzati, S. (2021). Unpacking algorithms as technologies of power: Syrian refugees and data experts on algorithmic governance. *Digital Geography and Society*, *2*, 100016.

Kaun, A. (2021). Suing the algorithm: the mundanization of automated decision-making in public services through litigation. *Information, Communication & Society*, 1–17

Kazansky, B., & Milan, S. (2021). "Bodies not templates": Contesting dominant algorithmic imaginaries. *New Media & Society*, *23*(2), 363–381. https://doi.org/10.1177/1461444820929316

Kennedy, H. (2016). Post, Mine, Repeat – Social Media Data Mining Becomes Ordinary. Palgrave Macmillan.

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.

Kerikmäe, T., Pärn-Lee, E. (2021) Legal dilemmas of Estonian artificial intelligence strategy: in between of e-society and global race. *AI & Soc 36*, 561–572. https://doi.org/10.1007/s00146-020-01009-8

Kerttunen, M. (2020, September 30). Küberturvalisus. Milline on meie küberstrateegia ja kas e-riik vajab remonti? Edasi – Innustav ja hariv ajakiri. Retrived from

https://edasi.org/66972/kuberturvalisus-milline-on-meie-kuberstrateegia-ja-kas-e-riik-vajab-remonti/
Keuschnigg, M., Lovsjö, N., & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, *1*(1), 3–14.
Kitchin, R. (2014). The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Consequences. London: Sage
Kivi, G. (2020, December 15). *nõusolekuteenus | Riigi Infosüsteemi Ameti blogi*. RIA Blogi. Retrived from https://blog.ria.ee/tag/nousolekuteenus/
Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information systems frontiers*, *19*(2), 267–283.
Klinger, U., & Svensson, J. (2018). The End of Media Logics? On Algorithms and Agency. *New Media & Society, 20(12)*, 4653–4670
Kõuts-Klemm, R., Pruulmann-Vengerfeldt, P., Siibak, A., Lauristin, M. (2017). Internetikasutus ja sotsiaalmeediakasutus. *Eesti ühiskond kiirenevas ajas: uuringu" Mina. Maailm. Meedia" 2002–2014 tulemused*. Vihalemm, P., et al., (eds). Tartu: Tartu Ülikool kirjastus, 279–298
Krippendorf, K. (2016). Data. In International Encyclopedia of Communication Theory and Philosophy. New York: Wiley. Retrived from https://onlinelibrary.wiley.com/doi/10.1002/9781118766804.wbiect104
Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data. Gartner Blog Network. Retrived from http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming- gartners-volume-velocity-variety-construct-for-big-data/
Layder, D., (2006) Understanding Social Theory. 2nd ed. London: Sage
Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. *Big Data & Society, 6(1)*, 1–12. https://doi.org/10.1177/2053951718821146
Lindgren, S. (2019). Hacking social science for the age of datafication. *Journal of Digital Social Research*, *1*(1), 1–9
Lomborg, S., & Kapsch, P. H. (2020). Decoding algorithms. *Media, Culture & Society*, *42*(5), 745–761.
Maciejewski, M. (2017). To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences*, *83*(1_suppl), 120-135.
Magalhães, J., & Couldry, N. (2021). Giving by Taking Away: Big Tech, Data Colonialism, and the Reconfiguration of Social Good. *International Journal Of Communication, 15 (20)*. Retrieved from https://ijoc.org/index.php/ijoc/article/view/15995/3322
Mager, A. (2018). Internet governance as joint effort: (Re)ordering search engines at the intersection of global and local cultures. *New Media & Society 20(10)*, 3657–3677
Mahrenbach, L. C., & Mayer, K. (2020). Framing policy visions of big data in emerging states. *Canadian Journal of Communication*, *45*(1), 129–141
Majandus- ja Kommunikatsiooniministeerium and Riigikantselei (2019). Eesti tehisintellekti kasutuselevõtu eksperdirühma aruanne. Retrived from https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne.pdf
Mansell, R. (2012). *Imagining the Internet: Communication, innovation, and governance*. Oxford University Press.
Masso, A., Lauristin, M. Opermann, S., Kalmus, V. (2020). Applying the morphogenetic perspective for the analysis of Estonian social transformations. (1–32). In:

Kalmus, V., Lauristin, M., Opermann, S., Vihalemm, T. (eds). *Researching Estonian transformation: Morphogenetic reflections*. Tartu: University of Tartu Press

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60–68.

McKelvey, F. (2014). Algorithmic media need democratic methods: Why publics matter. *Canadian Journal of Communication*, *39*(4), 597–613.

McNeely, C. L., & Hahm, J. (2014). The big (data) bang: Policy, prospects, and challenges. *Review of Policy Research, 31(4)*, 304–310.

Mejias, U. A., & Couldry, N. (2019). Consumption as Production. Data and the reproduction of capitalist relations. *In The Oxford Handbook of Consumption*, Wherry, F.F. & Woodward, I. (eds), Oxford Univeristy Press. pp. 269–280.

Merriam, S. B. (2009). Qualitative research: A guide to design and implementation. San Francisco, CA: John Wiley & Sons.

Metz, J. (2015). The European Commission, Expert Groups, and the Policy Process. Demystifying Technocratic Governance. European Administrative Governance Series. Basingstoke: Palgrave Macmillan

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)*. https://doi.org/10.1177/2053951720948087

Minevich, M. (2021, January 17). *How The EU Is Leading The Way In AI Powered Social Innovation*. Forbes. Retrived from https://www.forbes.com/sites/markminevich/2021/01/18/how-the-eu-is-leading-the-way-in-ai-powered-social-innovation/?sh=30b9d0406bb7

Mittelstadt, B. D., & Floridi, L. (2015). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics, 22(2)*, 303–341.

Mosco, V. (2015). *To the cloud: Big data in a turbulent world*. Routledge.

Murumaa-Mengel, M., Pruulmann-Vengerfeld, P., & Laas-Mikko, K. (2014). Privaatsusõigus inimõigusena ja igapäevatehnoloogiad. *2014. aasta Inimõiguste Instituudi uuring*.

Nagy, P., & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. *Social Media+ Society*, *1*(2), 2056305115603385.

National AI strategy for 2019–2021 gets a kick-off – e-Estonia. (2019, October 31). Retrieved from https://e-estonia.com/nationa-ai-strategy/

Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, *10* (17).

Noble, S.U. (2018). Algorithms of Oppression. How Search Engines Reinforce Racism. New York: New York University Press

Nyman Metcalf, K. (2019). How to build e-governance in a digital society: the case of Estonia. *Revista Catalana de Dret Públic, (58)*, 1–12. https://doi.org/10.2436/rcdp.i58.2019.3316

O'Neil, C. (2016). Weapons of Math Destruction. New York: Crown Publishing

Olbrich, P. & Witjes, N. (2016). Sociotechnical imaginaries of big data: commercial satellite imagery and its promise of speed and transparency. In: Bunnik A, Cawley A, Mulqueen, M, et al. (eds) *Big Data Challenges*. London: Palgrave Macmillan, 115–126

Paige, J. B., & Morin, K. H. (2016). Q-sample construction: A critical step for a Q-methodological study. *Western journal of nursing research*, *38*(1), 96–110.

Park, G.; Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Kosinski, M.; Stillwell, D. J.; … Seligman, M. E. P. (2015), 'Automatic personality assessment through social media language,' *Journal of Personality and Social Psychology, 108 (6)*, 934–952. https://doi.org/10.1037/pspp0000020

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Pink, S., Lanzeni, D. and Horst, H. (2018). Data anxieties: finding trust in everyday digital mess. *Big Data & Society, 5(1)*. DOI: 10.1177/2053951718756685

Plantera, F. (2017) "Artificial Intelligence is the next step for e-governance in Estonia." [Online]. Retrived from: https://e-estonia.com/artificial-intelligence-is-the-next-step-for-e-governance-state-adviser-reveals/

Plantera, F. (2019, December 5). *Data tracker - tool that builds trust in institutions*. E-Estonia. Retrived from: https://e-estonia.com/data-tracker-build-citizen-trust/

Prainsack, B. (2019). Logged out: Ownership, exclusion, and public value in the digital data and information commons. *Big Data & Society, 6(1),* DOI: 10.1177/ 205395171982977.

Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication*, *8*, 1690-1709.

Puusalu, S. (2020). SUURANDMED: OLEMUS JA KASUTAMISE KITSASKOHAD. Sisekaitseakadeemia. Retrived from: https://digiriiul.sisekaitse.ee/bitstream/handle/ 123456789/2537/2020%2007%20suurandmed-web.pdf?sequence=1&isAllowed=y

Quan-Haase, A. (2016). Technology & Society. Social networks, power, and inequality. Don Mills, Ontario: Oxford University Press

Redden, J. (2018) Democratic governance in an age of datafication: Lessons from mapping government discourses and practices, *Big Data & Society, 5 (2)*, 1–13

Redden, J., Dencik, L., & Warne, H. (2020). Datafied child welfare services: Unpacking politics, economics, and power. *Policy Studies*, *41*(5), 507–526.

Reddy, E., Cakici, B. and Ballestero, A. (2019.) Beyond mystery: Putting algorithmic accountability in context. *Big Data & Society 6(1)*, 2053951719826856

Resnyansky, L. (2019). Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. *Big Data & Society*, *6*(1), 2053951718823815.

Ribes, D., and Jackson, S.J. (2013). Data Bite Mam: The Work of Sustaining a Long-Term Study. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 147–166.

Rieder, G. (2018). Tracing big data imaginaries through public policy: the case of the European Commission. In *The Politics of Big Data* (pp. 89–109). Routledge.

Rieder, G., & Simon, J. (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society*, *3*(1), 2053951716649398. https://doi.org/10.1177/2053951716649398

Rogers, E. (2003). *Diffusion of Innovations*. Fifth edition. Free Press: New York.

Rössler, B. (2015). Should personal data be a tradable good? On the moral limits of markets in privacy. *Social dimensions of privacy: Interdisciplinary perspectives*, 141–161.

Roßnagel, A. (2020). Evaluation of the General Data Protection Regulation. *Data protection data security* 44, 287–292 https://doi.org/10.1007/s11623-020-1269-1

Ruppert, E. (2011). Population objects: Interpassive subjects. *Sociology 45(2),* 218–233

Ruppert, E. (2015) Doing the Transparent State: Open Government Data as Performance Indicators, in R. Rottenburg, S.E. Merry, S.J. Park, and J. Mugler (eds.), *A World of*

Indicators: The Making of Governmental Knowledge through Quantification, Cambridge, Cambridge University Press, pp. 127–150.
Ruppert, E. (2018). Sociotechnical imaginaries of different data futures: An experiment in citizen data. van Doorn leerstoel, Erasmus Universiteit Rotterdam. Retrieved from www.panart.nl
Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 2053951718820549.
Schäfer, M. T. (2011). Bastard Culture! User Participation and the Extension of Cultural Industries. Amsterdam: Amsterdam University Press.
Schäfer, M. T. & Clausen, N., (2021). Participatory Data Ethics. In: Wienrich, C., Wintersberger, P. & Weyers, B. (Hrsg.), Mensch und Computer 2021 – Workshopband. Bonn: Gesellschaft für Informatik e.V. DOI: 10.18420/muc2021-mci-ws06-316
Schmitz, A. J. (2014). Secret consumer scores and segmentations: Separating haves from have-nots. *Mich. St. L. Rev.*, 1411.
Scholl, R., Velsberg, O. (2020). AI FOR GOOD LIVE: How Estonia builds the next generation e-government with AI use cases. International Telecommunications Union. Retrived from: https://www.youtube.com/watch?v=O3jiLbYVLQU (Nov. 10, 2020)
Schroeder, R. (2014). Big data and the brave new world of social media research. *Big Data & Society, 1(2)*, 1–11
Selwyn, N., & Pangrazio, L. (2018). Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data & Society*, *5*(1), 2053951718765021.
Siegel, E. (2013). Predictive Analytics. Hoboken: Wiley
Silaškova, J., & Pappel, I. (2021). How COVID-19 has pushed Estonia towards next generation citizen services | *Baltic Rim Economies*. Retrieved from https://sites.utu.fi/bre/how-covid-19-has-pushed-estonia-towards-next-generation-citizen-services/
Solove, D. J. (2008). The end of privacy? *Scientific American*, *299*(3), 100–106.
Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, *54*(4), 208–216.
Stephenson, W. (1953). *The study of behaviour: Q-technique and its methodology*. Chicago: The University of Chicago Press.
Suzor, N. (2019). Lawless: The Secret Rules that Govern Our Digital Lives. Cambridge: Cambridge University Press
Svensson, J., & Poveda Guillen, O. (2020). What is Data and What Can it be Used For?: Key Questions in the Age of Burgeoning Data-essentialism. *Journal of Digital Social Research (JDSR)*, *2*(3), 65–83.
Tammpuu, P., & Masso, A. (2018). 'Welcome to the virtual state': Estonian e-residency and the digitalised state as a commodity. *European Journal of Cultural Studies, 21(5)*, 543–560.
Taylor, L., & Purtova, N. (2019). What is responsible and sustainable data science? *Big Data & Society, 6(2)*, 1–6. https://doi.org/10.1177/2053951719858114
Tene, O. & Polonetsky, J., (2013) Big data for all: Privacy and user control in the age of analytics. Retrived from http:// heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11&section=20
Thatcher, J. (2014). Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication, 8(1)*, 1765–1783.

Thatcher, J. E., & Dalton, C. M. (2022). Life in the Age of Big Data. In *Data Power: Radical Geographies of Control and Resistance* (pp. 13–45). Pluto Press. https://doi.org/10.2307/j.ctv249sg9w.8

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, *34*(6), 990–1006.

Thimm, C., Anastasiadis, M., & Einspänner-Pflock, J. (Eds.). (2017). *Media logic (s) revisited: Modeling the interplay between media institutions, media technology and societal change*. Jumper

Van den Boomen, M. (2014). Transcoding the digital. Amsterdam: Institute of Network Cultures.

Van der Vlist, F. N., & Helmond, A. (2021). How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society*, *8*(1), 20539517211025061.

Van Dijck J, Nieborg, D. and Poell, T. (2019) Reframing platform power. *Internet Policy Review 8(2)*, 1–18.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, *12*(2), 197–208.

Van Dijck, J. (2020a). Governing digital societies: Private platforms, public values. *Computer Law & Security Review*, *36*, 105377.

Van Dijck, J. (2020b). Seeing the forest for the trees: Visualizing platformization and its governance. *new media & society*, 1461444820940293.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

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.

Van Exel, J., & De Graaf, G. (2005). Q methodology: A sneak preview.

Vihma, P. (2020, October 14). *Estonia to digitise the world through the WHO*. Invest in Estonia. https://investinestonia.com/estonia-to-digitise-the-world-through-the-who/

Watts, S., & Stenner, P. (2012). Introducing Q methodology: the inverted factor technique. *Doing Q methodological research theory, Method & Interpretation. London, UK: SAGE Publications*, 3–23.

Williamson, B. (2015). Smarter learning software: Education and the big data imaginary. In *Big Data—Social Data*.

Winner, L. (1993). Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology, & Human Values, 18(3): 362–378*.

Woodruff, A & Fox, S E, Rousso-Schindler, S, & Warshaw, J (2018) 'A qualitative exploration of perceptions of algorithmic fairness', CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 2018 Paper No.:656. https://doi.org/10.1145/3173574.3174230

Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society*, *43*(5), 807–824.

Zerilli J, Knott A, Maclaurin J, et al. (2018) Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology*. https://doi.org/10.1007/s13347-018-0330-6

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs.

SUMMARY IN ESTONIAN

Andmepioneeride suurandmetega seotud kujutluspildid: muutunud andmesuhted ja agentsust puudutavad väljakutsed

Andmetest, eriti suurandmetest – ja aina enam ka tehisintellektist – räägitakse tihti kui millestki müstilisest. Samas on aga nii suurandmetel põhinevate analüüside kui tehisintellekti taga alati erinevad (andme)eksperdid, kellel on suur roll tänase digiühiskonna kujundamisel ja uurimisel. Minu doktoritöö keskmes on seega just inimeste, konkreetsemalt andmeekspertide roll suurandmete kasutamisel, andmetehnoloogiate arendamisel ning seeläbi ka lõppkasutaja kogemuse kujundamisel. Oma doktoritöös analüüsisin suurandmetega seotud kujutelmi. Uurisin, milliseid võimalusi seoses suurandmete kasutamisega andmeeksperdid näevad era-, avaliku- ja kolmanda sektori asutustes ning milliseid võimalikke probleeme kogevad andmete kasutamispraktikas. Täiendavalt uurisin, kuidas on teadlaste tööd ja praktikaid mõjutanud domineerivad suurandmetega seotud kujutelmad ning milliseid väljakutseid suurandmete kasutamine kaasa on toonud.

Püstitasin doktoritöös kaks uurimisküsimust, millele andsin uuringute tulemustele tuginedes ka järgmised vastused:

1) **Millised on andmepioneeride valitsevad arusaamad ja kujutlused seoses suurandmetega?** (**UURING II, III, IV**)
   a) Suurandmeid peetakse peamiselt väärtuslikuks ressursiks, mis pakub vahendeid valitsemise toetamiseks parema otsustusprotsessi ja tõhususe suurendamise kaudu. Andmeid käsitletakse üha enam kauba ja kapitalina. Neid peetakse oluliseks konkurentsieeliseks nii era- kui ka avaliku sektori organisatsioonide jaoks. Andmeekspertidega läbiviidud intervjuude põhjal võib see selge eelis osutuda aga koostöö tegemisel mõnel juhul ka takistuseks, kuna andmeanalüüside kaudu loodetakse andmetest leida ennekõike oma organisatsiooni jaoks kasulikke mustreid, mille põhjal paremaid otsuseid teha (**UURING III** ja **IV**).
   b) Lisaks kujutatakse andmeid, eriti sotsiaalmeedia suurandmeid, võimalusena paremini mõista sotsiaalseid protsesse ja uurida inimeste käitumist. Kuna sotsiaalmeedia suurandmete kasutamise algusaastatel ei olnud metoodiline tööriistakast ja oskused (eriti sotsiaalteadustes) alati piisavad, innustas see interdistsiplinaarset koostööd tegema ning erinevaid valdkondi koondavaid keskusi looma. Koostöö arvutuslikke meetodeid kasutavate teadlastega andis sageli paremaid võimalusi suurte andmemahtude kogumiseks ning analüüsimiseks. Lisaks erinevate distsipliinide koostööle, soodustas sotsiaalmeedia suurandmete parem kättesaadavus ka uute meetodite ja tööriistade väljatöötamist, ning nende igakülgset testimist. Eelkõige algusaastail olid teadlased innukad erinevatelt platvormidelt kogutud sotsiaalmeedia andmestikke läbi töötama ja kombineerima. Kuigi viimaste põhjal tehtud varajased uuringud aitasid kaasa metodoloogilistele arengutele ja sisuga seotud uuendustele, viitavad minu doktoritöö

leiud ka sellele, et epistemoloogilised uuendused tekivad alles pärast uute analüüsitehnikate ja -meetodite väljatöötamist ja rakendamist (**UURING II**).

c) Erinevate suurandmete kujutelmade realiseerimine osutus aga tihti mitmete takistuste tõttu keeruliseks. Näiteks käsitleti sageli andmepõhiste lähenemisviiside rakendamise peamise takistusena andmete juurdepääsuprobleemi. Uuringute põhjal olid probleemiks veel: suurandmete analüüsimise ebapiisavad oskused ja teadmised, andmete jagamiseks vajalike standardite puudumine, andmesubjekti õiguste kaitseks seotud õiguslikud piirangud, aga ka muutunud andmesuhted (**UURING II** ja **III, IV**).

d) Kuigi oskuste puudumist ja nende arendamist toetas erinevate valdkondade koostöö ja erinevad digimeetodite koolitused, on (sotsiaalmeedia) platvormide tehnoloogilised võimalused seadnud piirangud andmetele juurdepääsu osas. Kuna juurdepääs andmetele antakse pigem huvirühmadele, kes neid oma ärilistel eesmärkidel kasutavad, siis toob see esile ka andmesuhete muutuva dünaamika. Seega, on andmed ennekõike asutuste poolt kogutav ja nende vahel jagatav-müüdav ressurss, kus andmesubjekte koheldakse üha enam kui kaupa (**UURING II**).

e) Andmesuhete muutused mõjutavad ka andmeeksperte. Eriti on see tähelda tav kolmanda sektori organisatsioonides, kus sõltuvus klientide ootustest toob kaasa sellise andmepõhise lähenemise, mida pigem kujutavad ette kliendid, mitte aga organisatsioonides töötavad andmeeksperdid. Kuigi mõnel juhul võivad eksperdid aktiivselt püüda oma klientidele uusi lähenemisi tutvustada ja pakkuda, võivad tekkida erinevad infrastruktuuri puudutavad – aga ka õiguslikud – piirangud, mis omakorda erineval moel nende uute lähenemiste rakendamist piiravad (**UURING III, IV**).

2) **Milliseid väljakutseid esitab ühiskondlik andmestumine individuaalsele ja institutsionaalsele toimevõimele?** (**UURING I, II, III, IV**)

a) Eelnevalt välja toodud piirangud esitavad paratamatult ka väljakutseid nii individuaalsele kui ka institutsionaalsele agentsusele. Peamiselt mõjutavad ekspertide agentsust nõuded või piirangud seoses andmetehnoloogiate kasutamisega organisatsioonides. Eraettevõtteid juhib tavaliselt äriline imperatiiv, mis on ajendatud nii nende enda kui ka erinevate, sageli nähtamatute sidusrühmade eesmärkidest. Tulemused viitavad sellele, et andmeeksperdid ja sotsiaalmeedia suurandmete uurijad lähtuvad pigem väärtuse imperatiivist, kuna eesmärgiks on andmeanalüüsi kaudu klientide/ühiskonnale väärtust (uusi teadmisi) pakkuda. Mõlema rühma jaoks on aga järjest olulisem leida ka tasakaal klientide ja andmesubjektide vajaduste vahel (**UURING II, III** ja **IV**).

b) Andmeekspertide jaoks on andmed ühiskondade tõhususe, parema kontrolli ja sotsiaalse hüve peamine allikas. Eksperdid tajuvad endal väga suurt rolli selliste hüvede loomisel. Seetõttu juurutavad mõned neist oma organisatsioonides aktiivselt uut, andmepõhist kultuuri ja defineerivad end ka ise selgelt andmeentusiastidena. Sel põhjusel võiks eksperte näha

kui andmepioneeride kogukonda, kellel on aktiivne ja oluline roll andmepõhiste teadmiste vahendamisel nii oma organisatsioonis, klientidele, aga ka laiemalt ühiskonnas. (**UURING III**).

c) Üksikisikutel kui andmesubjektidel (antud töös käsitletud ka kui andmeloojatel) puudub, tuginedes andmeekspertide arusaamadele, andmepõhises ühiskonnas kontroll oma andmete üle. Kuigi digitaalne kirjaoskus on oluline, muudab teatud protsesside läbipaistmatus üha keerulisemaks ka institutsioonide poolt tehtud otsuste vaidlustamise. (**UURING III, IV**).
d) Uute lähenemisviiside tutvustamisel rõhutasid andmeeksperdid oma vastutust avalikkuse ees ning läbipaistvuse ja vastutuse ideaalide tähtsust. Need ideaalid on peamiselt seotud võimalusega andmete põhjal tehtud otsuseid vaidlustada. Andmeeksperdid rõhutasid organisatsioonide vastutuse olulisust, kuna üksikisikutel ei ole sageli selget ülevaadet, kes ja millistel eesmärkidel nende andmeid kasutab. Seetõttu ei võimalda hetkel kasutatavad meetmed (nt teavitamine ja nõusolek) sageli üksikisikutel kui andmesubjektidel omada kontrolli oma andmete üle ning seeläbi ka oma privaatsust kaitsta. See omakorda suurendab aga ka institutsionaalse usalduse tähtsust. (**UURING I** ja **III**).

PUBLICATIONS

CURRICULUM VITAE

Name: Maris Männiste
Date of birth: February 23, 1987
Phone: +372 54530089
E-mail: maris.manniste@ut.ee

Education:

2016–2022	University of Tartu, PhD, Media and Communication
2013–2015	University of Tartu, MA, Knowledge and Information Management
2006–2010	University of Tartu, BA, Journalism and Communication

Work experience:

2021–...	University of Tartu, Faculty of Social Sciences, Institute of Social Studies, Junior Lecturer in Information Systems (1.00)
2021–2022	University of Tartu, Faculty of Social Sciences, Institute of Social Studies, Junior Research Fellow
2019–...	University of Tartu, Faculty of Social Sciences, Institute of Social Studies, Analysis of Society and Information Processes Programme Manager
2017–...	University of Tartu, Faculty of Social Sciences, Institute of Social Studies, Information and Knowledge Management Master's Curricula Programme Manager
2015–2021	University of Tartu, Faculty of Social Sciences, Institute of Social Studies, Assistant in Information Systems

Publications:

Appelman, N., von Brakel, R., Calatayud, J. M., Chelioudakis, E., Chiusi, F., Izdebski, K., Kaun, A., Männiste, M., Fathaigh, R., Toh, J., Visser, C., Quevedo, W.C. (2021). Tracing The Tracers 2021 report: Automating COVID responses. A report coordinated by AlgorithmWatch (Germany).

Masso, A., Männiste, M & Siibak, A. (2020). 'End of Theory' in the Era of Big Data: Methodological Practices and Challenges in Social Media Studies. Acta Baltica Historiae et Philosophiae Scientiarum, 8 (1), 33−61. DOI: 10.11590/abhps.2020.1.02.

Männiste, M. & Masso, A. (2020). "Three Drops of Blood for the Devil": Data Pioneers as Intermediaries of Algorithmic Governance Ideals. Mediální studia / Media Studies, 14 (1), 55−74.

Männiste, M. (2020). Automated Decision Making in Estonia. Automating Society. A report coordinated by AlgorithmWatch (Germany).

Männiste, M. & Masso, A. (2018). The Role of Institutional Trust in Estonians' Privacy Concerns. Studies of Transition States and Societies, 10 (2), 22−39.

Conferences:

Masso, A. & Männiste, M. – End of Theory in the Area of Big Data: Methodological Practices and Challenges in the Social Media Studies. Data Power, 22.06.–23.06.2017, Carleton University, Ottawa, Canada.

Männiste, M. & Masso, A. – End of Theory in the Area of Big Data: Methodological Practices and Challenges in the Social Media Studies. Digital Culture Meets Data: Critical Approaches, 06.11.2017–07.11.2017, Brighton University, Brighton, England

Männiste, M. & Masso, A. – Datafication of Mobile live: discrimination through algorithmic selection. Data Power: Global In/securities, 12.09.2019 –13.09. 2019, Bremen, Germany.

Männiste, M. – Datafication of Mobile live: discrimination through algorithmic selection. ECREA Midterm conference "DATAFICATION, MEDIATIZATION AND THE MACHINE AGE", 01.11.2019–02.11.2019, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany.

Männiste, M. – WORKSHOP TOWARDS DEVELOPMENT OF MEDIATIZATION RESEARCH III, 15.11.2019–15.11.2019, Institute of Journalism and Social Communication, University of Wroclaw; Academia Europaea Wroclaw Knowledge Hub, Poland.

Teaching experience:

Lecturer: I have taught several courses concerning the basics of information systems. In recent years I've also taught courses more focused on critical data science, for example "Contemporary Issues of the Information Age" (2017–), and "Big data and the datafied society" (spring 2021 and spring 2022).

ELULOOKIRJELDUS

Nimi: Maris Männiste
Sünniaeg: 23. veebruar 1987
Telefon: +372 54530089
E-post: maris.manniste@ut.ee

Haridus:

2016–2022	Tartu Ülikool, PhD, meedia ja kommunikatsioon
2013–2015	Tartu Ülikool, MA, organisatsiooni infotöö
2006–2010	Tartu Ülikool, BA, ajakirjandus ja suhtekorraldus

Töökogemus:

2021–...	Tartu Ülikool, Sotsiaalteaduste valdkond, Ühiskonnateaduste instituut, infosüsteemide nooremlektor (1.00)
2021–2022	Tartu Ülikool, Sotsiaalteaduste valdkond, Ühiskonnateaduste instituut, andmestunud ühiskonna uuringute nooremteadur
2019–...	Tartu Ülikool, Sotsiaalteaduste valdkond, Ühiskonnateaduste instituut, ühiskonna ja infoprotsesside analüüsi õppekava programmijuht
2017–...	Tartu Ülikool, Sotsiaalteaduste valdkond, Ühiskonnateaduste instituut, info- ja teadmusjuhtimise õppekava programmijuht
2015–2021	Tartu Ülikool, Sotsiaalteaduste valdkond, Ühiskonnateaduste instituut, infosüsteemide assistent

Publikatsioonid:

Appelman, N; von Brakel, R.; Calatayud, J. M., Chelioudakis, E., Chiusi, F., Izdebski, K., Kaun, A., Männiste, M., Fathaigh, R., Toh, J., Visser, C., Quevedo, W.C. (2021). Tracing The Tracers 2021 report: Automating COVID responses. A report coordinated by Algorithm Watch (Germany).

Masso, A., Männiste, M. & Siibak, A. (2020). 'End of Theory' in the Era of Big Data: Methodological Practices and Challenges in Social Media Studies. Acta Baltica Historiae et Philosophiae Scientiarum, 8 (1), 33−61. DOI: 10.11590/abhps.2020.1.02.

Männiste, M. & Masso, A. (2020). "Three Drops of Blood for the Devil": Data Pioneers as Intermediaries of Algorithmic Governance Ideals. Mediální studia / Media Studies, 14 (1), 55−74.

Männiste, M. (2020). Automated Decision Making in Estonia. Automating Society. A report coordinated by AlgorithmWatch (Germany).

Männiste, M. & Masso, A. (2018). The Role of Institutional Trust in Estonians' Privacy Concerns. Studies of Transition States and Societies, 10 (2), 22−39.

Konverentsid:

Masso, A & Männiste, M. – 'End of Theory' in the Area of Big Data: Methodological Practices and Challenges in the Social Media Studies. Data Power, 22.06.2017–23.06. 2017, Carleton University, Ottawa, Kanada

Männiste, M. & Masso, A. – 'End of Theory' in the Area of Big Data: Methodological Practices and Challenges in the Social Media Studies. Digital Culture Meets Data: Critical Approaches, 06.11.2017–07.11 2017, Brighton University, Brighton, Inglismaa

Männiste, M. & Masso, A. – Datafication of Mobile live: discrimination through algorithmic selection. Data Power: Global In/securities, 12.–13. September 2019, Bremen, Saksamaa

Männiste, M. – Datafication of Mobile live: discrimination through algorithmic selection. ECREA Midterm konverents "DATAFICATION, MEDIATIZATION AND THE MACHINE AGE", 01.11.2019–02.11.2019, Rheinische Friedrich-Wilhelms-Universität Bonn, Saksamaa

Männiste, M. – WORKSHOP TOWARDS DEVELOPMENT OF MEDIATIZATION RESEARCH III, 15.11.2019–15.11.2019, Institute of Journalism and Social Communication, University of Wroclaw; Academia Europaea Wroclaw Knowledge Hub, Poola

Õpetamiskogemus:

Lektor: Olen õpetanud mitmeid infosüsteemide aluseid puudutavaid aineid, viimastel aastatel ka aineid, mille sisu keskendub ka andmestunud ühiskonna teooriatele ja uurimise meetoditele nagu Infoühiskonna aktuaalsed teemad (2017–) Suurandmed ja andmestunud ühiskond (2021 kevad, 2022 kevad).

DISSERTATIONES
DE MEDIIS ET COMMUNICATIONIBUS
UNIVERSITATIS TARTUENSIS

1. **Epp Lauk**. Historical and sociological perspectives on the development of Estonian journalism. Tartu, 1997, 184 p.
2. **Triin Vihalemm**. Formation of collective identity among Russophone population of Estonia. Tartu, 1999, 217 p.
3. **Margit Keller**. Representations of consumer culture in Post-Soviet Estonia. Tartu, 2004, 209 p.
4. **Pille Pruulmann-Vengerfeldt**. Information technology users and uses within the different layers of the information environment in Estonia. Tartu, 2006, 213 p.
5. **Anu Masso**. Constitution of personal social space in A transition society. Tartu, 2008, 212 p.
6. **Kristina Reinsalu**. The implementation of Internet democracy in Estonian local governments. Tartu, 2008, 176 p.
7. **Andra Siibak**. Self-presentation of the "Digital Generation" in Estonia. Tartu, 2009, 257 p.
8. **Pille Runnel**. The Transformation of the Internet Usage Practices in Estonia. Tartu, 2009, 223 p.
9. **Tõnu Tender**. Mitmekeelsus Eestis Euroopa Liidu mitmekeelsuse ideaali taustal. Tartu, 2010, 253 lk.
10. **Roosmarii Kurvits**. Eesti ajalehtede välimus 1806–2005. Tartu, 2010, 424 lk.
11. **Kadri Ugur**. Implementation of the concept of media education in the Estonian formal education system. Tartu, 2010, 153 p.
12. **Barbi Pilvre**. Naiste meediarepresentatsioon Eesti ajakirjanduskultuuri ja ühiskonna kontekstis. Tartu, 2011, 201 lk.
13. **Kertu Saks**. The formation of editing culture and practice in Estonian newspapers 1988–2005. Tartu, 2011, 161 p.
14. **Andres Jõesaar**. EU media policy and survival of public service broadcasting in Estonia 1994–2010. Tartu, 2011, 147 p.
15. **Tiiu Kreegipuu**. The ambivalent role of Estonian press in implementation of the Soviet totalitarian project. Tartu, 2011, 221 p.
16. **Indrek Treufeldt**. Ajakirjanduslik faktiloome erinevates ühiskondlikes tingimustes. Tartu, 2012, 290 lk.
17. **Maarja Siiner**. Towards a more flexible language policy: a comparative analysis of language policy design in Denmark and Estonia. Tartu, 2012, 231 p.
18. **Taavi Tatsi**. Transformations of museum-embedded cultural expertise. Tartu, 2013, 148 p.
19. **Krista Lepik**. Governmentality and cultural participation in Estonian public knowledge institutions. Tartu, 2013, 205 p.

20. **Lukas Blinka**. The "good" and the "bad" of the internet: Studying subjective well-being as an outcome of young people's online practices. Tartu, 2013, 171 p.
21. **Ragne Kõuts-Klemm**. Media-connected society: the analysis of media content and usage patterns from a systems-theoretical perspective. Tartu, 2013, 286 p.
22. **Maie Kiisel**. Problems of critical analysis of communication of environmental issues and risks. Tartu, 2013, 208 p.
23. **Marek Miil**. Nõukogude propagandasüsteemi toimimine ajakirjanduse argipraktikate kaudu. Tartu, 2014, 269 lk.
24. **Katrin Kello**. The functions and contexts of general education history teaching: social and professional representations in Estonia and Latvia. Tartu, 2014, 230 p.
25. **Maie Soll**. The ethnic identity of russian-speaking students in estonia in the context of educational change. Tartu, 2015, 212 p.
26. **Linda Lotina**. Conceptualizing engagement modes: Understanding museum – audience relationships in Latvian museums. Tartu, 2015, 186 p.
27. **Andres Kõnno**. The Modelling of Communication and Its Applications in a Longitudinal Research: Examples of Estonian, Finnish and Russian Journalism in the 20th Century. Tartu, 2016, 250 p.
28. **Olga Einasto**. Academic Library E-service Quality and Working User: Conceptual Model. Tartu, 2016, 186 p.
29. **Maria Murumaa-Mengel**. Managing Imagined Audiences Online: Audience Awareness as a Part of Social Media Literacies. Tartu, 2017, 184 p.
30. **Jan Richard Bærug**. The Collapsing Wall. Hybrid Journalism. A Comparative Study of Newspapers and Magazines in Eight Countries in Europe. Tartu, 2017, 301 p.
31. **Agnese Karaseva**. Teacher Professional Agency in Relation to Digital Technology Integration in Teaching in Estonian and Latvian Schools. Tartu, 2017, 164 p.
32. **Kadi Lubi**. The role of information search and interpretation in living with chronic illness. The case of Estonian Parkinson's disease patients. Tartu, 2017, 152 p.
33. **Marju Himma-Kadakas**. Skill performance of Estonian online journalists: assessment model for newsrooms and research. Tartu, 2018, 144 p.
34. **Johannes Saar**. Hegemooniadiskursused eesti kultuuris. Eesti Kunstimuuseumi pressikommunikatsioon 2006–2015. Tartu, 2018, 179 lk.
35. **Alessandro Nanì**. Cross-Media in Public Service Broadcasting: The Struggle Between Producers and Audiences. Tartu, 2018, 176 p.
36. **Signe Ivask**. The role of routines, demands and resources in work stress among Estonian journalists. Tartu, 2018, 138 p.
37. **Karmen Palts**. Lastevanemate ja õpetajate vaheline kommunikatsioon Eesti esimese kooliastme näitel. Tartu, 2019, 189 lk.
38. **Tiiu Männiste**. Elukestev õpe ühiskonna pöördeaegadel. Tartu, 2019, 301 lk.

39. **Mari-Liisa Parder**. Communication of Alcohol Consumption Practices and Situational Abstinence as a Basis of Prevention: A Study of Estonian Adolescents. Tartu, 2020, 152 p.
40. **Marianne Leppik**. The segmented integration and mediated transnationalism of Estonian Russian-speaking populations. Tartu, 2020, 279 p.
41. **Olena Nedozhogina**. Identity construction of Russian-speaking Ukrainians after 2013–2014. Tartu, 2021, 168 p.