



UNIVERSITY  
OF TURKU

# IMAGINING THE DATA ECONOMY

---

Tuukka Lehtiniemi





UNIVERSITY  
OF TURKU

# IMAGINING THE DATA ECONOMY

---

Tuukka Lehtiniemi

## University of Turku

---

Faculty of Social Sciences  
Department of Social Research  
Economic Sociology  
Doctoral Programme of Social and Behavioural Sciences

## Supervised by

---

Professor Pekka Räsänen  
Department of Social Research  
University of Turku

Associate Professor Minna Ruckenstein  
Centre for Consumer Society Research  
University of Helsinki

## Reviewed by

---

Professor Helen Kennedy  
Department of Sociological Studies  
University of Sheffield  
United Kingdom

Associate Professor Stefania Milan  
Department of Media Studies  
University of Amsterdam  
The Netherlands

## Opponent

---

Prof. Dr. Ingrid Schneider  
Department of Informatics  
University of Hamburg  
Germany

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Text © Tuukka Lehtiniemi. The synopsis part of this thesis is licenced under a CC BY-NC-ND 4.0 licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). The original publications are licenced under their respective CC licences.

ISBN 978-951-29-8001-7 (PRINT)  
ISBN 978-951-29-8002-4 (PDF)  
ISSN 0082-6987 (Print)  
ISSN 2343-3191 (Online)  
Painosalama Oy, Turku, Finland 2020

UNIVERSITY OF TURKU  
Faculty of Social Sciences  
Department of Social Research  
Economic Sociology  
TUUKKA LEHTINIEMI: Imagining the Data Economy  
Doctoral Dissertation, 190 pp.  
Doctoral Programme of Social and Behavioural Sciences  
April 2020

## ABSTRACT

The digital environment is increasingly organised to transform aspects of people's lives into data in order to make use of those data in the production of economic value. Data activism has emerged as one response to the resulting asymmetries in data usage and distribution. Adopting the concept of collective imagination, this thesis investigates imaginaries about an alternative data economy developed in data activism. Based on the data studies literature, a view of the dominant data economy imaginary is first constructed. It consists of collectively shared notions about how the data economy currently functions and ought to function. Based on four original publications, alternative imaginaries are compared with the dominant imaginary. The aim is to examine the alternative imaginaries and their underpinnings and to scrutinize the desirability of the data futures they promote.

Empirical research in this thesis has focused on MyData, a data governance initiative striving for a more central role for people in the data economy. The thesis identifies two alternative imaginaries developed in the context of the initiative: the market imaginary and the citizen imaginary. Both build on the notion of data agency, that is, providing people with new capabilities to act in relation to personal data. The market imaginary is based on viewing data agency as market choice, and relies on the market for data governance. Individuals are imagined to act in data markets to improve their lives, making data serve their personal ends. The citizen imaginary foregrounds collective data governance and the common good. Here, data agency is imagined as citizens' collective capability to participate in the processes that determine how and for what purposes their data are used.

This thesis discusses how the market imaginary is the better positioned of the two to expand beyond data activism; it resonates with notions about technology as the enabler of individual choice, leverages existing regulatory instruments, and is aligned with commercial views of the value of data. Based on this research, however, the reliance on market agency appears as a precarious starting point for a desirable data future. The practical implication is to encourage data activists to experiment on collective data governance and on new ways to make data valuable alongside the market-oriented ones. The implication for data activism research is that identifying imaginaries underpinning activist initiatives can aid with shaping pathways toward a desirable digital environment.

**KEYWORDS:** collective imagination, data activism, data agency, data citizenship, data economy, data governance, MyData, social imaginaries

TURUN YLIOPISTO

Yhteiskuntatieteellinen tiedekunta

Sosiaalitieteiden laitos

Taloussosiologia

TUUKKA LEHTINIEMI: Imagining the Data Economy

Väitöskirja, 190 s.

Yhteiskunta- ja käyttäytymistieteiden tohtoriohjelma

Huhtikuu 2020

## TIIVISTELMÄ

Digitaalisessa ympäristössä ihmisten elämästä tuotetaan dataa tavoitteena taloudellisen arvon luominen. Data-aktivismi on yksi reaktio datan käytön ja jakautumisen epäsymmetrioihin. Tässä työssä tarkastellaan kollektiivisen mielikuvituksen käsitteen avulla data-aktivismissa tuotettuja vaihtoehtoisia datatalouden mielikuvastoja, imaginaareja. Sosiaalitieteellisen datatutkimuksen avulla rakennetaan näkymä dominoivaan datatalouden imaginaariin. Tätä verrataan neljään tutkimusjulkaisuun pohjautuen data-aktivismissa tuotettuihin vaihtoehtoisiiin datatalouden imaginaareihin. Tavoitteena on paitsi eritellä vaihtoehtoisia imaginaareja, myös tarkastella niiden edistämien tulevaisuuden datatalouden muotojen tavoiteltavuutta.

Väitöskirjan empiirisenä tutkimuskohteena on MyData, datan hallintaa koskeva aloite, joka pyrkii muovaamaan ihmisille keskeisemmän rooliin tulevaisuuden datataloudessa. Tutkimuksen perusteella aloitteessa vaikuttaa kaksi datatalouden imaginaaria, jotka ovat sekä toisilleen että dominoivalle imaginaarille vaihtoehtoisia. Molemmat nojaavat ajatukseen ihmisten lisääntyvästä toimijuudesta suhteessa heitä koskevaan dataan. Markkinaimaginaari nojaa markkinoihin datan hallintamekanismina. Tällöin toimijuus kiteytyy valintoihin markkinoilla: ihmisten ajatellaan tekevän valintoja, joissa data palvelee heidän henkilökohtaista etuaan. Kansalaisimaginaari nojaa ajatukseen kollektiivisesta datan hallinnasta ja datasta saatavista yhteisistä hyödyistä. Toimijuus datan suhteen kytkeytyy tällöin kollektiivisiin prosesseihin, jotka ohjaavat datan käyttöä ja käytön päämääriä.

Tutkimuksen perusteella markkinaimaginaarilla on näistä kahdesta paremmat edellytykset laajentua data-aktivismiin ulkopuolelle: se hyödyntää olemassa olevaa sääntelyä, pohjautuu vakiintuneisiin käsityksiin datan taloudellisesta arvosta ja korostaa yksilön valinnanvapautta tukevaa teknologiaa. Tutkimuksen valossa luottaminen markkinatoimijuuteen yksinään vaikuttaa epävarmalta pohjalta tulevaisuuden datataloudelle. Työn käytännön implikaatio on rohkaista data-aktivismia kehittämään kollektiivisia datan hallinnan tapoja ja arvonmuodostuksen muotoja markkinavalinnan ohella. Työn implikaatio data-aktivismiin tutkimukselle on osoittaa, että data-aktivismiin taustalla vaikuttavien imaginaarien tunnistaminen auttaa kartoittamaan toivottavaa tulevaisuuden digitaalista ympäristöä.

ASIASANAT: data-aktivismi, datakansalaisuus, datan hallinta, datatalous, data-toimijuus, kollektiivinen mielikuvitus, MyData, sosiaaliset imaginaarit

# Acknowledgements

One figurative beginning for the path leading me to write this thesis was my very first day of school. I have a vivid memory of it; it had rained and I got my new white running shoes wet as I walked across the school lawn. It always begins with getting your feet wet. That schoolkid probably didn't think too much about the path ahead, but it now feels as if it was quite obvious all the way to the university. I lost the path for a while after grinding through a Master's in engineering and then in economics. As I worked outside of the academia, I day-dreamed about finding my way back in, if only to check if there was something I was missing out.

I eventually contacted Adjunct Professor Marko Turpeinen to see if he knew how a latecomer could get started with doctoral studies. It turned out he did, and more quickly than anyone would believe I had quit my job in favour of a research project and was tasked to get accepted in a doctoral programme. Professor Pekka Räsänen agreed to supervise my thesis in economic sociology even if my background was outside of the field. Thanks to Marko and Pekka for taking me onboard at your respective institutions, for the open-mindedness and respect towards my ideas, for smoothing out the bureaucracy, and for the important feedback and suggestions I have received. I would not be here without your help.

I am grateful that Associate Professor Minna Ruckenstein agreed to act as the co-supervisor. It must have been an experience for an anthropologist to supervise an engineer-economist who wants to be an economic sociologist, but I don't think anyone could have done it better. Apart from providing guidance supervisors also set the bar, and I learned that once you thought my work was good enough, it really was good enough. This thesis would be quite different without your help.

Professor Helen Kennedy and Associate Professor Stefania Milan kindly acted as the reviewers of this thesis. I am grateful and honoured that two leading experts in the field have evaluated my work. I thank you for the encouraging feedback, and for pointing out how to make the thesis stronger. I am also very pleased to have Prof. Dr. Ingrid Schneider as my opponent in the public examination.

To paraphrase one famous author, I don't know if easy articles are all alike since I'm yet to write one, but every difficult article surely is difficult in its own way. I

thank Jesse Haapoja and Dr. Yki Kortensniemi for shovelling academic snow with me. I have learned a lot from working with you.

I have been in an exceptional position to do research on collective imagination thanks to two creative individuals, Dr. Kai Kuikkaniemi and Antti Jogi Poikola. Thank you for all the masterminding. I also thank all of my interlocutors in the MyData community and beyond for their input and contributions, and the MyData community for the open-mindedness towards us who stick to the outskirts. I have tried to give in return, I hope some of it is helpful.

Peer review is an integral but often derided part of the academic life. I may have been lucky, because in my experience peer review only improves the papers. I thank all journal editors and anonymous reviewers who have provided feedback on my work. I especially thank one editor for an early remark on social imaginaries, which eventually led me to discover key analytical concepts for the thesis.

During these years I have had four academic homes: Digital Content Communities research group at Aalto University, the discipline of Economic sociology at University of Turku, Centre for Consumer Society Research at University of Helsinki and the scientific association Rajapinta. Having four homes means I have enjoyed the wisdom, support and great company of more colleagues than I can keep track on. For me, writing is the deepest form of thinking, and people who have helped me think by reading and commenting my manuscripts include at least Dr. Vassilis Charitsis, Dr. Titiana Ertiö, Dr. Nina Janasik, Dr. Mikko Jauho, Ilkka Koiranen, Dr. Aki Koivula, Dr. Salla-Maaria Laaksonen, Dr. Airi Lampinen, Dr. Ella Lillqvist, Dr. Matti Nelimarkka, Professor Mika Pantzar, Dr. Olli Pitkänen, Pasi Pohjalainen, Juho Pääkkönen, Dr. Henrik Rydenfelt, Dr. Arttu Saarinen, Dr. Outi Sarpila, Sonja Trifuljesko and Dr. Julia Velkova. I have enjoyed the insights and company of also Kirsikka Grön, Dr. Ville Harjunen, Dr. Jukka Huhtamäki, Dr. Kai Huotari, Arto Kekkonen, Dr. Jens Kremer, Mea Lakso, Petri Lievonen, Robin Lybeck, Viivi Lähteenoja, Mimmi Piikkilä, Dr. Essi Pöyry, Anna Rantasila, Dr. Mikko Rask, Jukka Reitmaa, Dr. Petteri Repo, Minna Saariketo, Tuomas Soila, Maiju Tanninen, Minttu Tikka, Dr. Päivi Timonen, Dr. Kalle Toiskallio, Pihla Toivanen, Dr. Linda Turunen, Dr. Antti Ukkonen, Teppo Valtonen, Sanna Vellava and Marjoriikka Ylisiurua.

My 2018 fellowship at Alexander von Humboldt Institute for Internet and Society in Berlin was made productive and memorable by Dr. Stefan Baack, Edoardo Celeste, Dr. Christian Djefal, Dr. Benedikt Fecher, Dr. Alexandra Giannopoulou, Kirsten Gollatz, Dr. Astrid Mager, Philip Meier, Shirley Ogolla, Christopher Olk, Martin Riedl, Dr. Max von Grafenstein and the whole HIIG community.

Outside of the academic world, I have had influential exchanges on the topic of this thesis with Daniel Kaplan, Valérie Peugeot and Molly Schwartz.



This work would not have been possible without the financial support from Tekes, Helsingin Sanomat Foundation, Kone Foundation, KAUTE foundation and Turku Center for Welfare Research. My sincere thanks to these institutions and the people working behind the curtains.

Climbing has been a perfect complement to working on a thesis, and both seem to require a taste for the long haul. I thank Dr. Jussi Heinonen for the countless hours at the gym and the crag, Dr. Otto Kässi for the company in the mountains, and all other climbing friends for sharing the rope with me.

In my childhood home, I grew up to value knowledge, education, informed opinion and self-reflection. This has been a solid foundation to build on. Thanks to my mother for always being genuinely interested in the details of what I'm doing. It's much rarer that you'd think. I believe I know what my late father would say: I've done a didit! Thanks to my sister and brothers for making me feel I'm appreciated.

The last and the greatest thanks go to my wonderful wife Salla-Maaria. When you worked on your own thesis, I saw how committed and content you were, and this had a great deal to do with why I started working on mine. Our relationship has always been both romantic and intellectual, and being colleagues has been an occasionally weird but always wonderful addition to it.

Now I know there was something in the academia that I was missing out. Doing a PhD has been the most challenging, rewarding, intensive and stressful period in my life. I'm grateful that I did it, and I'm grateful that it's done.

Hermanni, Helsinki, 25 February 2020  
*Tuukka Lehtiniemi*



## **TUUKKA LEHTINIEMI**

Trained first as an electrical and communications engineer (M.Sc.) and then as an economist (M.Soc.Sc.), Lehtiniemi became interested in how the uses we invent for new technologies are shaped by how we imagine the economy to work. Besides the academic life, he enjoys other precarious situations such as steep uphill, high places and balancing on hands.

# Preface

It can be quite difficult to pinpoint when an idea, such as an idea for a thesis, emerged. Luckily datafication, which I will properly define later, can help. Researchers not only meticulously document their own work by means of notes, annotations, memos and research plans. Nowadays, mundane parts of the research process also automatically produce metadata. Reference management software shows when an article was discovered, and time stamps of annotations reveal when it was read. The forensics made possible by these data traces made it surprisingly easy to track down when the questions addressed in this thesis started taking form.

I began working on this thesis in the summer and early autumn of 2014 at the Helsinki Institute for Information Technology when I joined a multi-disciplinary research consortium doing research on data and knowledge work. I had already been exposed to celebratory accounts on online platforms and big data (e.g., Anderson, 2008; Mayer-Schönberger & Cukier, 2013). While reading for the project, I encountered more critical literature on the topic. Based on Mendeley time stamps, these included work by Mark Andrejevic (2014), danah boyd and Kate Crawford (2012), Tarleton Gillespie (2010) and José van Dijck (2014).

I was also affected by the Finnish discussion on individuals' rights to their "digital footprint" (e.g., Kuittinen & Ruckenstein, 2014; Pitkänen, 2014). Around that time, a working group of Finnish open data proponents, some of whom were my new colleagues, prepared a report for a Finnish Ministry (Poikola et al., 2014). It described MyData, a data governance concept aiming at what was (and is) imagined as a more sustainable, human-centric digital environment built on the understanding that people, firms, and society at large would benefit if individuals were equipped with means to control their personal data and would thus become more active consumers and participants.

As reflected in the first iterations of the research plan for this thesis, I was interested in what I considered the personal data economy – that is, in conceptualising the collection and use of personal data in terms of economic activity, including value creation, production, exchange, consumption, prosumption and distribution. Inspired by the data studies literature, discussions on individuals' rights in the digital environment, and new models for organising the production and use of

personal data, initial research questions started taking form. Are people becoming economic actors toward their data in some new sense? What does that imply for the data economy? What could participation in the data economy mean? In addition to the Finnish MyData initiative, do others have similar aims? Later, after a more thorough plunge into the data studies literature, more critical questions began emerging: Is this something that we should strive for? Paraphrasing the title of boyd and Crawford's influential 2012 article, what are the critical questions for my data?

This thesis gathers together five years of work that started with these questions. The original research is reported in peer-reviewed and published articles, and this book serves as an introduction and contextualisation of that research, a summary of its findings, and a means to draw the findings together to answer a set of broader research questions. In addition to datafication and data economy, key concepts of this thesis include data activism and collective imagination. The emerging stream of research on data activism helped to conceptualise MyData and similar initiatives as data activism in the context of the data economy. The literature on collective imagination offered an overall analytical framework to connect the original publications. The title of the thesis is based on a small trope in the literature on collective imagination and owes to Robin Mansell's 2012 book *Imagining the Internet*, which also has an overall framework somewhat similar to mine.

A research aim motivating this work was to better understand the empirical field of data activism. I was from the beginning both sympathetic and critical toward the data activists' aims and approaches. I strived to engage with the field in a manner that would be committed to a critical scholarly position but that could simultaneously be constructive and offer something for data activism. This resulted in a research approach of taking different engagement positions in relation to the field. This turned out to be a fruitful research strategy, even if at times it necessarily became a balancing act.

A final note on the nature of this work – the starting point of data activism examined in this thesis is to make better, not to start from scratch. They are not the activists who take to the figurative streets of the digital environment to smash data capitalism; rather, they intervene in the prevailing data arrangements, aiming to make them serve their own ends. This thesis is about critically examining both the data economy and data activism, aiming to produce scholarly knowledge as well as to improve data activism (which, in turn, tries to improve the data economy). This means that the thesis, like the data activists it examines, does not smash but rather tries to assemble; it operates “within,” hopefully helping to intervene and improve.

# Table of Contents

<b>Acknowledgements</b> .....	<b>v</b>
<b>Preface</b> .....	<b>viii</b>
<b>Table of Contents</b> .....	<b>x</b>
<b>List of Original Publications</b> .....	<b>xiii</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 The data economy .....	2
1.2 Datafication .....	4
1.3 Datafication as a societal transformation .....	6
1.4 Data and capitalism.....	8
1.5 Data activism .....	11
1.6 Collective imagination .....	12
1.7 Research aims and research questions .....	14
1.8 The empirical context .....	14
1.9 The structure of the thesis.....	17
<b>2 Collective Imagination</b> .....	<b>18</b>
2.1 Collective imagination in social theory .....	19
2.2 Collective imagination and technology .....	20
2.3 Future imaginaries and economic dynamics.....	23
2.4 Collective imagination as an analytical tool .....	24
<b>3 The Dominant Data Economy Imaginary</b> .....	<b>27</b>
3.1 The economic arguments for datafication.....	28
3.2 Datafication and informational capitalism .....	31
3.3 The social relations of data extraction .....	34
3.4 Formal institutions regulating data extraction .....	36
3.5 Resignation towards dataveillance .....	39
3.6 The data economy imaginary .....	40
<b>4 Data Activism</b> .....	<b>43</b>
4.1 Civic engagement and the digital environment .....	44
4.2 Data activism as a heuristic tool .....	46
4.3 Data activism and social justice.....	47
4.4 Varieties of data activism .....	48
4.5 Data activism and the private sector.....	51

4.6	MyData as data activism .....	52
<b>5</b>	<b>Research Approach .....</b>	<b>56</b>
5.1	Research approach of the original publications .....	58
5.2	Research approach of the thesis .....	64
<b>6</b>	<b>Findings .....</b>	<b>70</b>
6.1	Article I: Consent intermediaries .....	70
6.2	Article II: Personal data spaces .....	73
6.3	Article III: Data agency at stake .....	75
6.4	Article IV: The social imaginaries of data activism .....	78
<b>7</b>	<b>Discussion .....</b>	<b>81</b>
7.1	Data agency and the politics of imagination.....	82
7.2	The market imaginary.....	85
7.3	The citizen imaginary .....	88
7.4	The success of alternative imaginaries.....	90
7.5	Desirable data futures .....	92
<b>8</b>	<b>Conclusion .....</b>	<b>97</b>
	<b>References .....</b>	<b>101</b>
	<b>Original Publications .....</b>	<b>113</b>



# List of Original Publications

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

- I Lehtiniemi, T., & Kortensniemi, Y. (2017). Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach. *Big Data & Society*, 4(2), 1–11. <https://doi.org/10.1177/2053951717721935>. Licenced under CC BY-NC.
- II Lehtiniemi, T. (2017). Personal data spaces: An intervention in surveillance capitalism? *Surveillance & Society*, 15(5), 626–639. <https://doi.org/10.24908/ss.v15i5.6424>. Licenced under CC BY-NC-ND.
- III Lehtiniemi, T., & Haapoja, J. (2020). Data agency at stake: MyData activism and alternative frames of equal participation. *New Media & Society*, 22(1), 87–104. <https://doi.org/10.1177/1461444819861955>. Licenced under CC BY-NC.
- IV Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. *Big Data & Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718821146>. Licenced under CC BY-NC.

The original publications have been reproduced under the terms of their respective Creative Commons licences.





# 1 Introduction

The digital environment is increasingly organised to transform aspects of people's lives into data, and the purpose of producing data is often to make use of those data in the production of economic value. This has created an inherently asymmetric situation in terms of data usage and distribution. Some actors, such as digital platform companies, are well-positioned and well-equipped to produce and use data. Other actors, such as small firms or the public sector, might be less capable of making use of data. People whom the data concern fundamentally lack the means to make data serve their own ends.

This thesis investigates data activism that has emerged as a response to the asymmetries of data usage and distribution in the context of the data economy. The focus is on imaginaries about the data economy, that is, collectively held notions about how the data economy works and ought to work. The thesis compares alternative data economy imaginaries developed in data activism with the dominant imaginary about the data economy and with each other.

Our data future remains uncertain and unknowable. How the future is imagined to unfold and how current actions are imagined to affect the future nevertheless make a difference, as they have consequences in the present. The dominant imaginary about the data economy affects the development of data technologies and the data arrangements in which they are embedded. Compared to this, data activism, its imaginaries about an alternative data economy, and its alternative data infrastructures and data arrangements represent a possibility for change. This is to say that imaginaries about the societal and economic outcomes of new data technologies affect the technologies that are developed, delimit how they are used, and shape the data arrangements in which these technologies and their use are embedded. Imaginaries have implications, making data economy imaginaries – both dominant and alternative – a relevant research topic.

This thesis consists of a synopsis part and four original research publications. This introductory chapter starts with describing the broader context of the data economy and datafication and then proceeds to introduce other central concepts of the thesis: data activism and collective imagination. It also lays out the research approach, research questions, and the empirical context of the thesis.

## 1.1 The data economy

Celebratory and more critical accounts of the contemporary digital economy agree on at least one issue; the creation of economic value relies heavily on the unprecedented visibility, knowability and calculability of actors and their behaviours and preferences. The digital economy's new means of value creation are made possible by the transformation of aspects of people's lives into quantified data and then turning those data into information and knowledge. Data on our everyday lives including consumption, work, health, mobility, communication, social interaction, and leisure, are increasingly captured, stored in databases, processed, analysed or mined, shared, assessed, and acted on in various ways by the state as well as commercial actors (Kitchin, 2014). Technological advancements such as digital platforms, sensors, internet connectivity of devices in personal, domestic, and industrial contexts, and tracking techniques online and in the physical world make possible the production of more and more data. In 2012, an industry estimate claimed that 90% of all existing data was produced within the two previous years, and around the same time it was estimated that the production of data will continue to increase around 40% annually (see Kennedy, 2018). It is difficult to fathom the volume of all data that humanity and its devices currently produce. A recent estimate puts the amount of data produced each day at 2.5 quintillion bytes (Marr, 2018), and it is estimated that in 2020 there will be 40 times as many bytes of data as there are stars in observable universe (Domo, 2019).

This thesis concerns the economic system consisting of the production, distribution, exchange and utilisation of data by and about people. By way of a definition of what I am discussing, I will use the term *personal data economy* or, interchangeably, *data economy* to refer to the activities providing goods and services that serve human needs based on personal data. The data economy is here understood in terms of its function – activities intended to meet human needs (Elder-Vass, 2016: 28–29; 2018). Understood in this way, the data economy does not refer only to production that meets market demand. While it includes such production, it also covers things that are provided without charge as well as things that fulfil the need for public goods. The focus is particularly on economic activities that concern *personal data* or, as they are sometimes called, user data or human subject data. I consider personal data as any data related to or resulting from actions by a person. More colloquially, personal data refers to data by and about people.

For the discussion in this thesis, this definition of the data economy serves the purpose of simply delineating as the object of interest actions that produce goods and services that serve human needs and that are based on personal data. For more practical purposes such as producing measures and statistics about the data economy, the definition could lead to difficulties that would call for more specificity. These difficulties could concern both the data and the economy. For example, it might not

always be obvious whether data are related to an individual and should therefore be considered personal data. As another example of potential difficulties, specifying economic activity in terms of provisioning departs from definitions underlying macroeconomic statistical aggregates, which are typically defined in terms of market exchange.

Nevertheless, numerical estimates of the economic value produced in the context of the data economy give some sense of the scale of the phenomenon. A 2013 consultancy estimate puts the potential annual economic value of all digitally available information about individuals in Europe only at around 1,000 billion euros in 2020 (Boston Consulting Group, 2013). This figure attempts to include the economic value of these data to all parties involved, including organisations and consumers; in other words, it does not include just market activities, and it does not measure the value of data in the context of any currently measurable aggregate figure such as the gross domestic product. According to another estimate that specifically involves macroeconomic aggregates, the value of the European data economy in 2017 was 335.5 billion euros, or 2.4% of the EU GDP (Cattaneo et al., 2018: 15). This figure attempts to include “the direct, indirect, and induced effects of the data market on the economy as a whole” (Cattaneo et al., 2018: 45) – that is, the contribution of both personal and non-personal data to the EU GDP. These figures and the reports describing the processes that go into producing them are, among other things, testimonials on the conceptual and practical difficulties of measuring the economic value of data.

Another way to provide concrete context for the production of economic value involving personal data is to look at the data-producing and data-using companies and their revenue streams. At the end of 2018, the fourth-largest public corporation in the world, as measured by market capitalisation, was Google’s parent company Alphabet, and the sixth-largest was Facebook (Statista, 2019c). The annual revenues of Google and Facebook in 2018 were 136 billion and 56 billion USD, respectively. Advertising brought in around 86% of Google’s revenue and more than 98% of Facebook’s revenue (Statista, 2019a; 2019b). These firms have reached their positions in the advertising market by means of collecting personal data via commercial surveillance on their platforms and elsewhere, turning those data into predictions of users’ behaviour and interests and targeting advertisements according to these predictions (Rieder & Sire, 2014; West, 2019). Advertising revenues are therefore intimately connected to, and made possible by, personal data. To be clear, these figures contain the value of those personal data-based commodities that are exchanged on the market. They do not measure the economic value of data only; a number of other things in addition to data go into the production of value including inputs such as computational and storage resources, programme code, intellectual capital, and human labour as well as the creation of necessary market arrangements.

As a summary, the data economy and other key terms employed throughout this thesis are listed in Table 1.

**Table 1.** Summary of key terms employed in the thesis.

<b>Term</b>	<b>Definition</b>
Collective imagination	Imagination as social phenomenon. Produces imaginaries, or collective and normative beliefs of how the society functions and ought to function. Holds together communities and institutions but also produces visions of counterfactual futures and alternative social realities (Beckert, 2016; Taylor, 2004; Jasanoff, 2015a).
Data activism	Civic engagement and political action responding to the uneven distribution of data and data-related capabilities in datafied times (Baack, 2015; Kennedy, 2018; Milan & Gutiérrez, 2015; Milan & van der Velden, 2016).
Data agency	Capability to act intentionally in relation to the production and use of personal data.
Data arrangements	The practices of producing data, making use of data, and exploiting data for economic value as well as the policies governing these practices.
Datafication	The transformation of aspects of life into quantified data by means of information technology (van Dijck, 2014).
Personal data	Any data related to an individual, or resulting from actions of an individual. Data by and about people.
(Personal) data economy	The economic system that provides goods and services for human needs based on personal data.
MyData	A data governance initiative aiming at a more central role for people in the digital environment based on the understanding that people, companies, the public sector and society at large benefit if individuals control personal data (MyData Global, 2019; Poikola et al., 2015).

## 1.2 Datafication

The emergence of the data economy can be viewed through two related but conceptually distinct phenomena: digitalisation and datafication (Flyverbom et al., 2019). In the conventional sense of the term, digital refers to anything involving information technology (Peters, 2016), and digitalisation refers to the employment of information technology to translate processes, inventories and records into digital format. Digitalisation allows for storing more information as well as processing

information more efficiently and can bring obvious performance enhancements to various organisational, commercial and societal processes. Digitalisation has also contributed to many aspects of changes often associated with globalisation; with it, “phenomena such as outsourcing, the deterritorialisation of production, and the emergence of virtual, amorphous value chains became possible and advantageous” (Flyverbom et al., 2019: 6).

Rather than simply digitalisation of previously analogue things, the data economy is to a large extent the outcome of the other digital transformation, datafication. It is an outcome of a distinct capacity of information technology; in addition to simply automating things, it can create data on whatever it automates. One early exploration of this capacity was Shoshana Zuboff’s (1985, 1988) research on computer-mediated work in the 1980s. Zuboff conceptualised the data production capacity as the ability to “informate” and the outcome of informing as a comprehensive “textualisation” of the workplace, leading to the production of “the electronic text.” Datafication as a term is usually credited to Viktor Mayer-Schönberger and Kenneth Cukier (2013: 78), who lamented the lack of a term for the contemporary development allowing the quantification of various things that had so far been largely qualitative. The authors proposed the term, stating that “to datafy a phenomenon is to put it in a quantified format so it can be tabulated and analysed.” José van Dijck (2014) introduced the term in the social scientific context of data studies by framing datafication as something that has been gradually normalised into a new paradigm for understanding sociality and social behaviour. For van Dijck, datafication refers to “the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis” (2014: 198).

In practical terms, datafication now concerns a wide array of aspects of our lives. In the data economy, datafied aspects of people’s lives are raw materials to be extracted, mined for valuable insights, and turned into products. In social media and elsewhere, datafication turns friendship, emotional responses and interests into processable “algorithmic relations” (Bucher, 2012; van Dijck, 2014). It affects practices of clinical healthcare and self-care (Ruckenstein & Schüll, 2017), permeates the fields of online and physical retail (Turow et al., 2015b), makes possible personal profiling by means of location and mobility (Bellovin et al., 2013), and – through the internet of things and smart device development – makes possible the analysis of banal and mundane activities such as toasting the bread or opening the fridge.

From a technical perspective, datafication is driven by developments resulting in the capacity to produce and store data automatically. This can happen without the exertion of much effort once the automated process is set up and without action or even awareness from those being datafied. It is claimed to be easier to keep data that has once been produced and stored than it is to discard them (Mayer-Schönberger &

Cukier, 2013: 101). The technical ease of producing and storing data leads to concerns over the quality, efficiency, reliability, validity and usability of the means of data production, storage, sharing, processing and inference (Kitchin, 2014: 12–13). Attempts to ensure that human or technical deficiencies do not result in unreliable data or analysis results include systematic data collection and analysis procedures, equipment specification, technical compensation of biases and errors in analysis, or the adherence to technical schemas or standards (Kitchin, 2014). A focus on technical aspects of datafication is closely related to the idea that this development is unavoidable but needs quality control.

### 1.3 Datafication as a societal transformation

To paraphrase Kranzberg's (1986) first law of technology, while not necessarily a force for good or ill, datafication is not neutral either. A key insight of surveillance studies (e.g., Lyon, 2007) can be borrowed here: surveillance does not simply relate to the mechanisms of surveillance but should rather be considered in the context of institutions and practices that enable these mechanisms and the ideological justification and normalisation of surveillance. Analogically, the same concerns datafication. It does not simply relate to technical capabilities of datafying. While datafication is made possible by technical capacity, it has gradually become “a trend advanced by the amalgamation of different cultural, political and economic forces that both shift and entrench power relations in contemporary society” (Hintz et al., 2019: 49).

Rather than a technical phenomenon related to the practical means of turning qualitative aspects of life into data, datafication is most usefully considered as a societal phenomenon. Following Rob Kitchin (2014: 12), considering datafication's societal significance requires framing it, in addition to technical terms, also in ethical, philosophical, political and economic terms. Such a broader perspective on datafication starts from the observation that the “data revolution” does not unfold in a passive, neutral and non-ideological manner but is rather driven forward by a powerful combination of normative arguments, the involved actors' beliefs, and economic and political interests (Kitchin, 2014). Datafication converts aspects of life to data, and the societally significant questions about datafication concern who makes use of those data, for what purpose, and what shapes the answers to these questions.

Following Kitchin, an ethical perspective on datafication concerns the right and wrong conduct in data production, sharing, analysis and use – in short, the extent to which data practices can be morally justified and under which circumstances (e.g., Herschel & Miori, 2017; Richards & King, 2014; Zwitter, 2014). Three branches of data ethics can be distinguished (Floridi & Taddeo, 2016). The ethics of data is

related to data generation, recording, curation, sharing and use. The ethics of algorithms is related to the increasing automation and complexity of processes employed in data collection, analysis, interpretation and decision-making. The ethics of practices is related to the questions of responsibilities and liabilities of people and organisations involved in the above processes and the devising of practices and professional code attempting to guide us toward ethical conduct.

A philosophical perspective to datafication focuses on the epistemological and ontological assumptions related to data and their use. Here, Kitchin (2014) identifies two endpoints of a spectrum. One is to consider data as something collected from the world in neutral and objective ways within technical constraints and subject to quality concerns. In this view, data are pre-analytical, pre-factual, and non-ideological; only the use of data can be political. The other is to view data as never “raw” but always “cooked” (Gitelman & Jackson, 2013) so that data are always a product of the thought systems and instruments that produce them (Bowker & Star, 1999). Contextual and material factors shape what is data and what data are. The transformation of human experiences into numerical values in databases for the purpose of learning via computational means produces particular forms of information and knowledge (Berry, 2011: 2; Hintz et al., 2019: 47). Emphasising the quantitative and computational means of understanding the social world “constitutes a novel, powerful system of knowledge with its own epistemology” (Milan & van der Velden, 2016: 63), delineating the limits of what can be known and affecting the way people relate to information and knowledge.

A political-economic perspective on datafication was apparent already in the 1980s explorations of informing the workplace, which pointed out how the introduction of information technology could produce radically divergent outcomes (Zuboff, 1985; 1988) depending on choices made regarding who can use new information to learn, how, and for what purposes. The outcome could be the amplification of the worst features of automation including the centralisation of authority and the depriving of workers of their autonomy; alternatively, workers could be empowered by the new information-creating capacity, engage in new kinds of learning processes and become an organisation’s most precious resources. More broadly, Arne Hintz and colleagues (Hintz et al., 2019: 4) point out that while the collection of data about people has long been at the heart of governance and control, questions about data are becoming ever more central in the contemporary society through the advancement and expansion of technological means for data production. The important issue here is not that databases contain more data or that these data are more valuable than before – rather, the important part is the new uses to which the data are put (Andrejevic & Gates, 2014: 186). Data, their production, and their interpretation are shaped by (and shape) the context in which they happen (Dalton & Thatcher, 2014) and thereby need to be examined as questions of power (Tufekci,

2014). Framing data, or access to data, as a form of power makes data an increasingly important issue in contemporary politics (e.g., Iliadis & Russo, 2016; Milan & Gutiérrez, 2015). The political-economic logic underpinning datafication matters because it affects the ways that society makes and can make use of data. The logic therefore shapes the societal transformations brought along with datafication; it “produces its own social relations and with that its conceptions and uses of authority and power” (Zuboff, 2015: 77).

## 1.4 Data and capitalism

Understanding datafication as a societal phenomenon motivates looking at it more broadly than a matter related only to technical capabilities and their development. This notion is in line with the view of the data economy that I am pursuing in this thesis. Sociological views of the economy have, since at least Max Weber, been based on the notion that economies consist not only of exchanges between economic actors but also of the social, cultural or institutional context in which those exchanges take place (Hass, 2007). Acknowledging this context means that production and exchange taking place in the data economy are viewed as culturally, institutionally and politically embedded. The cultural context affects what is understood as legitimate and normal economic practice; institutions such as property, contracts, markets and data protection shape and constitute economic action through rules and the associated roles of interaction, and political relations of power affect the distribution of material and economic resources (Hass, 2007: 16).

In the context of the data economy, scholars working on the political-economic aspect of datafication have outlined this broader context by examining key features running through the ways that data are used for economic profit (Andrejevic, 2014; 2016; Couldry & Mejias, 2019; Fourcade & Healy, 2017; Pasquale, 2017; Sadowski, 2019; Srnicek, 2017; Thatcher et al., 2016; van Dijck, 2014; West, 2019; Zuboff, 2015; 2019). The issues of concern include the normative construing of data as a legitimate object of economic interest such as a resource or an asset to own or trade; the understanding and analysis of the instituted economic logics and imperatives at work in our digital environment; the normative assumptions about normal, proper and expected data arrangements; the nature of data itself as a form of economic capital; and the assessment of data’s power relations. Related research also investigates the potential counteracting of these power differentials by means of governance and regulation of data and their use (e.g., Engels, 2016; Evans, 2017; Prainsack, 2019; Schneider, 2018).

One of the concepts employed to make sense of these developments is surveillance capitalism (Zuboff, 2015, 2019; on the previous use of the concept, see Foster & McChesney, 2004). Shoshana Zuboff describes surveillance capitalism as



an emergent economic form gaining power and hegemony in the data economy. In surveillance capitalism, firms aim to produce data about individuals for the purposes of knowing, controlling, and modifying behaviour, which are in turn employed to produce new varieties of commodification, monetisation, and control (Zuboff, 2015: 85). Its features include treating data as a resource for the taking, users as targets of data extraction through data-based surveillance, and extracted data as a form of capital. Data capital are then employed for automatic behavioural prediction and modification. To maximise the efficacy of their operations, many firms operate what economists call multi-sided markets (Rochet & Tirole, 2003), offering often free services to users and basing revenue streams on selling prediction-based products and services to paying customers who typically do not include the users. In Zuboff's analysis, the normative assumptions underlying the surveillance capitalism model – or the features of its economic logic – are viewed as being embedded in the ways in which data about people are collected, stored, and used.

How did this economic model come to be? A simple answer is that it emerged in the normal course of the evolution of capitalism. Citing Fernand Braudel (1984), Thomas Piketty (2014) and Roberto Unger (2007), Zuboff (2015) argues that there is no single variety of capitalism in terms of how production is organised; capitalism is “institutionally indeterminate” (Unger, 2007: 8), and the market economy can take variable forms of organisation of production, property and ownership, only small parts of which are realised at any given time (Piketty, 2014: 483). In fact, “capitalism's success over the *longue durée* has depended upon the emergence of new market forms expressing new logics of accumulation that are more successful at meeting the ever-evolving needs of populations and their expression in the changing nature of demand” (Zuboff, 2015: 77; also see Braudel, 1984: 620). Whereas this indicates continuous innovation on economic models over time, the market economy also seems to gravitate toward dominant models for value creation, which might become the institutionalised, taken-for-granted context in which companies operate. The production and consumption system known as Fordism is an example; its logic based on mass production of identical products and their mass consumption by the production workers took hold of industries and whole societies for decades, becoming the institutionalised, taken-for-granted context in which not only firms but also the society at large operated.

The company pioneering the economic model reliant on the production and use of personal data is, by many accounts, Google (e.g., West, 2019; Zuboff, 2015; 2019). Its economic model co-evolved with the search engines, as Google successfully created not only search algorithms but also a new business model with the aid of a specific combination of innovation culture and relatively lax data protection regulations (e.g., Elmer, 2004; Mager, 2012; 2017; van Couvering, 2008). The innovation that made Google immensely successful was to tap into the cache of

data that the firm had accumulated about its website users (West, 2019; Zuboff, 2019). Some of these data had been used to improve the search service provided to end-users, and the rest had been considered a surplus. Now this surplus was employed to predict the users' future behaviour by determining the probability that a specific individual would find a particular ad or offer interesting. Google then created auction-based markets for targeting services based on these predictions. As we now know, this contemporary form of "fortune telling and selling" (Zuboff, 2019: 88) turned out to be an immensely profitable endeavour. Google's position toward its users and customers, resulting from the combination of search and targeting, led to incentives to modify user behaviour in ways that serve the firm's own commercial ends (Rieder & Sire, 2014). Sarah Myers West (2019) argues that this model of operation was further strengthened by successfully masking the collection and use of personal data with a rhetoric of user empowerment and enhanced means of societal participation that came with the services offered. Previously, for example, José van Dijck and David Nieborg (2009) identified a similar business rhetoric on the benefits of democratized, transparent and collectivist digital space. Google's method of data collection about users was made possible by the institutional context of privacy protection, which acts as a countermeasure against some forms of surveillance but also as an enabler of extensive commercial data production (Coll, 2014; Draper & Turow, 2019). I will return to these aspects of the data economy and the institutional context in more detail in Chapter 3.

The significance of this economic model is not just that some large firms exhibit similar data arrangements. While the economic logic embodied in the data economy's successful exemplars is a relatively recent innovation, the argument made by Zuboff and others is that it has become an institution in itself, a model to which various kinds of firms looking to leverage data default. For example, competitive pressures push physical retail firms to mimic the surveillance practices of successful online firms (Turow et al., 2015b). More generally, the economic model seems to be mimicked by established firms in other sectors as well as up-and-coming start-ups in the form of institutional isomorphism – that is, "a constraining process that forces one unit in a population to resemble other units that face the same set of environmental conditions" (DiMaggio & Powell, 1983: 149). In this isomorphism of economic models, institutionalised pressures induce conformity to a specific economic logic that becomes the default model for data arrangements. Mechanisms of isomorphic change include coercive isomorphism through pressures exerted on organisations by other organisations on which they are dependent and by cultural expectations in the society as well as mimetic isomorphism resulting from voluntary imitation of models that are deemed acceptable or working, and they can be resorted to when facing uncertainties and poorly understood conditions (DiMaggio & Powell, 1983). In the context of isomorphic forms of economic utilisation of personal data,

coercive pressures might include funders demanding firms to gather data and employ them for value creation in a specific way, platform operators' coercive power over firms attached to them (Caplan & boyd, 2018), necessities introduced by the sheer scale of data-collecting operations (Andrejevic, 2016), and discursive strategies making it difficult to contest established economic expectations (e.g., Kitchin, 2014: 113; van Dijck, 2014; Sadowski, 2019). Mimetic pressures, on the other hand, could include the lure of lucrative profit opportunities exemplified by surveillance capitalism's pioneers.

## 1.5 Data activism

While capitalism's adaptability and institutional indeterminacy have led to one particular economic logic emerging and taking hold, they also mean that further innovation on novel data arrangements will continue to occur. Innovation on new data arrangements and practices no doubt currently happens in the context of commercial business development. At the same time, the cat-and-mouse play of commercial innovation and formal regulation continues as expected. State actors and policymakers have recognised the need to adapt old approaches and to develop new ones to engage with challenges and societally undesirable aspects of the data economy, slowly but nevertheless gradually aiming the regulatory apparatus at the economic models developed in the digital domain (Schneider, 2018).

Rather than commercial innovation or formal regulation, the focus of this thesis is on the civil society's inventiveness in responding to the commercial data arrangements. In particular, I will consider data activism as the context in which the underpinnings for alternative forms of the data economy's logic are being developed. Nevertheless, as will be discussed later, both commercial and regulatory developments have an important role to play; the form of data activism investigated here leverages both in order to reach its aims.

Data activism broadly refers to civic engagement and political action responding to the uneven distribution of data access and data-related capabilities in datafied times (Baack, 2015; Milan & van der Velden, 2016). Data activism, therefore, recognises the power differentials underlying the data economy and, in some sense, aims to affect or revert data power relations. Different initiatives have been conceptualised as data activism in the literature including employing data or data collection and analysis methods to improve the activists' or their interest group's condition, lobbying and implementing new ideas of data governance, or preventing data collection by means of technical tools and alternative platforms. For the purposes of this thesis, I follow Milan and van der Velden (2016) and consider data activism as a heuristic to examine engagement, action and participation that take a critical stand in relation to, but not necessarily in opposition to, datafication.

Data activists may or may not be explicitly interested in the data economy's economic logic; they might rather seek to employ the capacities of data technology to promote new forms of agency and participation, to mobilise data in order to enhance social justice, or to work against existing practices such as tracking or surveillance (e.g., Article III, Article IV; Baack, 2015; Gutiérrez, 2018a; Hintz et al., 2019; Kennedy, 2018; Milan & van der Velden, 2016). Nevertheless, when data activists experiment with and develop ways to act in relation to data, they engage in the political economy of data and channel discontent with the dominant industry practices.

The alternatives promoted and developed in data activism are obviously not mainstream; however, the novel data arrangements developed by data activists can nevertheless be significant. Building on previous scholarship on data activism (Baack, 2018a; Gutiérrez & Milan, 2019), data activists can be viewed not only as representing active engagement and political action with respect to datafication but also as a “front line” of developing alternative forms for the data economy through new data practices. Data activists “pioneer the exploitation of data infrastructure as a catalyst for social action” (Gutiérrez & Milan, 2019: n.p.), and such pioneer work expands the horizon of possibility to which others can orient themselves (Hepp, 2016). Therefore, data activism can represent the potential for meaningful change in the broader data arrangements (Schrock, 2016). The existing literature on data activism is discussed in more detail in Chapter 4.

## 1.6 Collective imagination

The line of inquiry I pursue in this thesis is based on the notion of collective imagination. The focus is on new, collectively imagined economic forms that introduce new pathways for the data economy. I will examine data activism as a context in which such new and potentially alternative pathways are explored and developed. Here, I build on the notion that data activists aim to develop alternative forms of collective imagination around datafication “to create a new sense for the legitimacy of collective knowledge creation” (Baack, 2015: 8).

The established concepts for discussing collective imagination in this sense are the social and sociotechnical imaginaries. Whereas the semantics of the imaginary in its contemporary everyday use involves the fancy, the non-reality, or the non-actuality, the concept is employed in a much different sense in philosophy, psychoanalysis, and – significantly for the purposes of this thesis – social theory and science and technology studies (see McNeil et al., 2017). In Charles Taylor's (2002, 2004) socio-political theory, social imaginary is the common understanding, shared by groups of people, of how the world works and what is normal. This ethos makes it possible for people to make sense of the society around them; it is constituted of

widely shared understandings that are taken for granted and that have achieved general legitimacy. In the context of examining scientific and technological progress, the notion of collective imagination is invoked in a more future-oriented manner, with a specifically articulated role for the material aspect of social order (i.e., technology). Sheila Jasanoff (2015a; also see Jasanoff & Kim, 2009) discusses the sociotechnical imaginary as a collectively held set of ideas, beliefs, and normative visions of desirable futures that are attainable through advances in science and technology. In both socio-political theory and studies of technoscience, imaginaries are considered as understandings that are taken for granted and, as a result, work implicitly in the background.

The approach in this thesis is inspired by the notion of multiple imaginaries. Even if imaginaries are collectively held, multiple imaginaries can nevertheless coexist in a society, be it in tension with one another or in a more productive, dialectical relationship (Jasanoff, 2015a: 4). Among the imaginaries, some prevail and emerge as dominant, becoming embedded in how actors in the society generally operate (Jasanoff & Kim, 2009: 123). Others could be considered “counter-imaginaries” (Hess, 2014), potential alternatives to the dominant ones. Heeding to these notions, I employ the concept of collective imagination in two senses. The first sense of employing collective imagination is to outline *that-which-is*, to depict the dominant imaginary of the data economy. This is the collective imagination that currently underpins the data economy – how the data economy is imagined to function, how personal data are understood to be normally exploited for the purpose of creating economic value and profit, and what role is consequently assigned to and assumed by people as participants in this economy. However, this dominant imaginary is not the only one; concurrent imaginaries can exist in the society, sometimes in a dialectical relationship, sometimes at odds with each other (Jasanoff, 2015a). This points to the second sense in which I will employ collective imagination: to outline *that-which-should-be*, to depict alternative data economy imaginaries. The alternative imaginaries developed in data activism contain the activists’ visions about a more desirable data future and different ways of functioning for the data economy. However, there is no reason to assume that data activism would be based on a uniform imaginary about the data economy. Taking the notion of multiple imaginaries to a logical end, one can begin by assuming that multiple concurrent imaginaries can also exist in data activism. Research on these alternatives makes it possible to do research on alternative possible futures (Jasanoff, 2015b: 339) contained within data activism. My interest in the dominant data economy imaginary is to use it in a comparative manner to analyse different alternative imaginaries. The literature on collective imagination and how collective imagination can be employed as an analytical tool is discussed in Chapter 2

## 1.7 Research aims and research questions

The original research done for this thesis is reported in the attached peer-reviewed and published articles I–IV. Each original publication has its own research questions and approach, collectively contributing to two broad research aims of this thesis. The first aim is to understand what the alternative data economy imaginaries developed in data activism are about and to pick apart their political and ideological underpinnings. This research aim informs the following two central research questions:

1. How do alternative data economy imaginaries and the dominant imaginary compare with each other?
2. How do different alternative imaginaries developed in data activism compare with each other?

The second research aim stems from the first one. If data activism produces not one but multiple alternative imaginaries, one logical follow-up question is whether some of these alternatives are more desirable than others. The second aim is to produce, through the understanding gained about the empirical phenomena, a position on what alternative imaginaries about the data economy could be considered to promote desirable data futures. This aim informs the third central research question:

3. How can we identify and promote societally desirable data economy imaginaries?

The research questions are developed into more specific sub-questions in Chapter 5.

## 1.8 The empirical context

As an empirical case of data activism, I examine MyData, a data governance initiative based on the idea that people should be able to exercise more control over personal data about them. Research for this thesis was to a large extent carried out in the context of participating in MyData-related research projects since 2014. MyData is originally a Finnish initiative aiming at a more sustainable digital environment based on the understanding that people, companies, the public sector and society at large benefit if individuals control the gathering, sharing and use of personal data (MyData Global, 2019; Poikola et al., 2015). The initiative is focused on redistributing data, particularly the benefits of data use, from data-gathering organisations to people. According to a white paper, the initiative’s goal is to turn people into “empowered actors, not passive targets, in the management of their personal lives both online and offline” (Poikola et al., 2015: 2). The present section provides a brief overview of MyData. Two of the original publications, Articles III and IV, consider MyData as a data activism initiative, focus on it as an empirical

case, and discuss the initiative and its origins more closely. Articles I and II focus on examining end-user data management services. There are better-known firms that promise to put organisations in charge of their data for their own advantage (Beer, 2018; Degli Esposti, 2014); the developers of the kinds of services examined in Articles I and II aim to do the same for people. They are closely related to MyData in that they have shared political aims with it, but they do not necessarily explicitly commit to a specific brand of data activism.

MyData originated with an Open Knowledge Finland working group, where it was initially developed in an open and collaborative manner by a group of open data activists and researchers along with businesspeople and civil servants, some of whom had more than one of these roles. This background invites one to consider MyData as a translation of open data ideas into the context of personal data. Whereas open data activists usually argue that data produced by public authorities should be technically and legally free for anyone to use, distribute and reuse (Kitchin, 2014), MyData translates the scope of both the data and the beneficiaries from the collective to the individual level. According to the MyData initiative, the right to decide on the uses of personal data collected by organisations should reside with the data subjects instead of being monopolised by the organisations. This does not mean that the initiative would frown upon commercial use of data as such; in contrast, a white paper portrays the initiative as aiming to provide the society with “parallel development of digital rights, innovation and business growth” (Poikola et al., 2015: 4). This can be interpreted so that MyData attempts to foster processes and policies for advancing individuals’ rights while concurrently accommodating the industry’s demands to process personal data in the production of economic and commercial value and in the development of new services. MyData seeks to simultaneously achieve these potentially divergent outcomes by rearranging the infrastructure underlying individual-level data practices; the newly imagined and gradually developed infrastructure comprises of personal data storages, data schemas and standards, exchange protocols, digital identity frameworks, and permission management tools. The private sector’s role is significant. MyData is built upon the understanding that to have an effect on the digital environment at large, the principles of individual data control developed in MyData need to be implemented by services developed by private firms.

Since the start of the research for this thesis in 2014, MyData has expanded from a working group of a Finnish open data NGO into an international non-profit organisation in its own right, MyData Global (2019), with local hubs listed in all continents. Nevertheless, as the initiative originated in Finland, it is in many ways grounded in the local institutional context and its embedded values. The above-mentioned white paper even explicitly frames the initiative as “a Nordic model for human-centric personal data management” (Poikola et al., 2015). One example of

this specific context is the early involvement of state actors. The white paper and its earlier Finnish-language version were commissioned by a Finnish government ministry. In addition to the socio-cultural context, the initiative is rooted in the formal regulatory environment. To achieve its aims, MyData leverages jurisdiction-specific regulations such as the General Data Protection Regulation in the EU, particularly its data portability rulings. Setting up an international NGO with local hubs speaks of attempts to transfer MyData's principles from the Nordic to other international contexts. In a different sense of transferability of the initiative's principles, individual data control is intended to be general, sector-independent, and ready to be embedded in any field-specific initiatives.

The idea of putting people in better control of their data is, of course, not new; it has been proposed under many labels in different times and contexts since at least the 1990s. One possibility to make sense of the discussions involved is to identify two strands based on whether the focus is on controlling data to limit adverse consequences of data production and use or on controlling data to make new data uses possible for the benefit of individuals (Iemma, 2016). Examples of the former include discussions in the context of ubiquitous computing environments (Bellotti & Sellen, 1993), a proposal for a market-based mechanism through which consumers could restrict the distribution of data by paying companies that have collected them (Noam, 1995), and more recent debates under the concept of privacy by design (e.g., Belli et al., 2017). Examples of the latter include the idea of gaining fair compensation for personal data use through regulated data markets (Laudon, 1996) and "infomediaries" envisioned as bargaining agents acting between consumers and businesses and making it possible for consumers to gain useful services in exchange for data and for companies to access a broad array of consumer data (Hagel & Rayport, 1997). In this pre-internet technological context, data ownership, paired with an intermediary facilitating data exchanges, was expected to shift the power balance toward consumers.

More recent developments in the same field include the "re-decentralisation" initiative by the web pioneer Tim Berners-Lee aiming to make personal data a resource for people (Andrejevic, 2014; Brooker, 2018), the development of software (Article II) or devices (e.g., Crabtree et al., 2016) to provide users with means to exercise control over data collection and use, and government-initiated "smart disclosure" programs releasing machine-readable personal data from firms to consumer-citizens (Iemma, 2016). Alex Pentland's (2009) "new deal on data," including among other things the notion that individuals should have the right to control their data, was introduced in a report by the World Economic Forum, which later declared personal data as a "new asset class" that individuals should control and benefit from (World Economic Forum, 2011). In a somewhat different take of the matter, Jaron Lanier proposes working toward commercial symmetry between firms



and users to be achieved by remunerating people whenever personal data are used (Lanier, 2013). A later iteration of the idea proposes treating data as labour and consequently introducing “data labour unions” (Arrieta Ibarra et al., 2018) or “mediators of individual data” (Lanier & Weyl, 2018) to take care of negotiations on the remunerations and terms of data use. By the end of 2019, data control and ownership had become mainstream enough to be minor talking points for would-be US presidential candidates (Levy, 2019; Molla & Stewart, 2019).

While these ideas have emerged elsewhere, many of them have affected MyData-related discussions, and some have resonated well with MyData proponents. The point here is not to investigate these similarities and differences further but to highlight that ideas about individuals’ capabilities to act in the digital environment and in relation to their personal data have been and still are under discussion, experimentation, debate and contestation in many locations and contexts. This connects the empirical investigations of this thesis to the broader theme of citizen agency and participation in datafied times.

## 1.9 The structure of the thesis

The remainder of the synopsis part of this thesis is organised as follows. The next three chapters examine the strands of research literature needed to reach the research aims. Chapter 2 establishes the use of collective imagination as an analytical tool, particularly as a tool of comparison between alternative imaginaries. Building on the data studies literature, Chapter 3 depicts a view of the dominant imaginary of the data economy that will provide a comparison point for the alternative imaginaries developed in data activism. Chapter 4 describes data activism as a field in which alternative imaginaries are being developed, discusses the variable forms of engagement with data and data technologies that data activism can include, and situates MyData in the data activism context. The remaining chapters lay out the original contribution of this thesis. Chapter 5 develops the research questions, presents the research approach, and discusses methods and data. Chapter 6 summarises the findings of the original research publications in light of the research aims of the thesis. Chapter 7 discusses the findings, and Chapter 8 concludes the thesis by summarising the results, pointing out their implications and outlining opportunities for further research in the area.

## 2 Collective Imagination

As Sally Wyatt points out, “the future has to be discussed in terms of the imaginary [...] but sometimes today’s imaginary becomes tomorrow’s lived reality” (Wyatt, 2004: 244). As it is discussed in this thesis, developing alternative imaginaries about the data economy means reimagining how the data economy functions, and what people’s role is in it – as citizens, consumers, or users. I do not refer to imagination as just a faculty of one person, whether a creative individual or a genius inventor. While an alternative data economy imaginary could be identified to originate in dreams of some key individuals who see beyond the constraints of what currently is, if an alternative data economy is to be realised, it will be “substantiated into people, objects, and practices” (Jasanoff, 2015b: 324) in a sense that surpasses individual creative faculties. The capacity to imagine is not only a capacity of individual human beings; rather, imagination is also a social phenomenon, and operates in an intersubjective, shared and collective manner (e.g., Beckert, 2016; Jasanoff, 2015a; Taylor, 2004). Imagination in the collective sense produces visions of counterfactual futures, alternative social realities, or new practices and arrangements; “it is through acts of imagination by collectives that start from somewhere different, not with solutions to problems already defined, but through practices of invention and experimentation that different futures can be performed” (Ruppert, 2018: 33).

The notion of collective imagination has inspired approaches from several disciplinary standpoints. It has its origins in a number of intellectual traditions in social theory, where collective imagination is considered as that which holds together large things: society’s institutions, market economies, national identities and modernity. Here, imagination concerns what is, as well as how it ought to be. While this notion of collective imagination is mainly oriented at the present, the concept has also been employed in an explicitly future-oriented manner in, for example, studies of technoscience and in examinations of capitalism’s success as a system of economic organisation. In this chapter, I discuss the literature dealing with collective imagination in the fields of political philosophy, science and technology studies and economic sociology. The aim is to establish the use of collective imagination as an analytical tool to address the research questions of this thesis. The resulting analytical approach will be to investigate the politics of collective

imagination by means of a comparative approach, distinguishing between a dominant data economy imaginary and its alternatives developed in data activism.

## 2.1 Collective imagination in social theory

In social theory, collective imagination refers to collective and normative beliefs, often implicit or tacit, of how the society functions and ought to function. It is the common understanding, shared by groups of people, of how the world functions and what is normal, enabling common practices and a shared sense of legitimacy. A key concept is the *social imaginary*, often employed in reference to the philosopher Charles Taylor and his work on modern social imaginaries (Taylor, 2002; 2004). However, before Taylor, other authors have written on collective imagination as well, employing varied terminology (see Adams et al. 2015; Flichy, 2007; Ruppert, 2018; Jasanoff, 2015a). Examples include the political philosophy on ideology and utopia by the philosopher Paul Ricœur (1986), the work on society's institutions by the philosopher, economist and psychoanalyst Cornelius Castoriadis (1987), as well as the work on nations as imagined communities by the political scientist Benedict Anderson (1991).

In terms of Ricœur's (1986) work, collective imagination (or "social and cultural imagination", in Ricœur's vocabulary) is a product of two contrasting components: ideological and utopian thought (Adams et al, 2015; Flichy, 2007: 8). Ideological thought is what maintains social order, legitimising and reproducing society's image of itself. Utopian thought is subversive and questions and problematises this order and produces alternative images for the society. Where ideology legitimises power, utopia provides an imagination for an alternative to power. Because of the interplay of ideology and utopia, there is a constant tension between the stability of the instituted social order, and its change. For Castoriadis (1987), collective imagination (or "social imaginary significations") is a way of understanding society's institutions; imaginaries are what underpin the institutions and the interwoven social practices of, for example, democracy, bureaucracy, or capitalism. To put it differently, for Castoriadis, these institutions are manifestations of a broader imaginary that is central to the existence of the society, so that society itself may be considered an imaginary institution (Adams et al., 2015). While imaginaries in this sense are necessary for society's institutions, there's no guarantee about their functioning and effect; in Castoriadis' view, then, while imaginaries are required for the functioning of institutions, they do not determine institutions (Ruppert, 2018). Anderson (1991) engaged with collective imagination in his work on nationalism and "imagined communities". For Anderson, the role of imagination is to hold together a nation; a heterogeneous and dispersed community that is made up of

people who may never encounter in face-to-face reality but nevertheless share, through collective imagination, practices which tie them together.

Taylor's work on social imaginaries (2002, 2004) is concerned with the question of how modernity came about. For Taylor, the notion of modernity includes new practices and institutional forms including technology, industrial production and urbanisation; the new ways of living such as individualism and secularisation; and new afflictions such as alienation and the experience of life's meaninglessness. In Taylor's account, central to the emergence of modernity is a new moral order of society, which came to influence the social imaginary that acts as the underpinning of modernity. Modernity, in short, came about as the social imaginaries underlying the society and its institutions and practices changed. For Taylor, social imaginary is the understanding, shared by groups of people, of how the world works and what is normal. It is that which "enables, though making sense of, the practices of a society" (Taylor, 2004: 2) and the "common understanding that makes possible common practices and a widely shared sense of legitimacy" (Taylor, 2004: 23). When understood in Taylor's sense, social imaginaries "define the contours of the social world and influence its governance" (Mansell, 2012: 33). Institutions such as the market and the civil society exist in the imaginations of actors who at the same time recognise that these notions are shared among other actors, which enables common practices around them (Kelty, 2008: 41–42). When common understandings are widely shared and taken for granted, they achieve general legitimacy.

The social imaginary, then, is something broader than a set of ideas or an intellectual scheme about social reality; it rather refers to "the ways people imagine their social existence, how they fit together with others, how things go on between them and their fellows, the expectations that are normally met, and the deeper normative notions and images that underlie these expectations" (Taylor, 2004: 23). The normative dimension of social imaginaries refers not only to the norms that structure our actions, but also the normative notions and expectations, as well as assumptions about expectations, of how the world usually works. According to Taylor, the shared understandings are "both factual and normative; that is, we have a sense of how things usually go, but this is interwoven with an idea of how they ought to go, of what missteps would invalidate the practice" (Taylor, 2002: 106). The foundations of social order, then, are made up of not only collective understanding of how the society functions, but also beliefs about how it is *supposed* to function.

## 2.2 Collective imagination and technology

In light of the research interests in this thesis, it is of particular interest that in the digital age, collective imagination influences both the development and the use of

digital technologies, and how they permeate and affect people's lives (Mansell, 2012: 33). In the context of technoscientific advancement, collective imagination has been invoked to account for implicit normative ideas underlying technology, such as collective notions of the correct role of technology in achieving societal and economic transformations. Collective imagination is a way to account for the normative aspect of technological advancement, that is, how technoscientific development is affected by a shared understanding of how things usually work and what is expected, and the normative notions that underlie these expectations. One example of such use of the concept is Kelty's (2008) ethnographic work on Free Software. Kelty builds on Taylor's discussion on social imaginaries and particularly their relationship to practices. In Kelty's account, Free Software practices are altered and transformed, or modulated, as they spread; as a result of these modulations, practitioners do not necessarily share the same goals, but what they do share is a social imaginary that "defines a particular relationship between technology, organs or governance [...] and the internet" (Kelty, 2008: 12). The point here is that decisions made about technology are embedded in imaginations of moral and technical order; social change enabled or facilitated by technology can encompass both ideas about the proper collective ordering of the economy and society, as well as the correct role of technologies in achieving that ordering.

In science and technology studies literature, collective imagination is employed in an explicitly future-oriented manner, so that collective imagination does not concern only to the role and use of available and existing technologies; rather, imaginaries project the future as it ought to be and contain a positive vision of technologically mediated progress. The key concept in this literature is the *sociotechnical imaginary*, developed particularly by the science and technology studies scholar Sheila Jasanoff (Jasanoff & Kim, 2009; Jasanoff, 2015a). Jasanoff takes the above-discussed social-theoretical accounts of the social imaginary as a starting point, but notes that while they consider imaginaries as things that shape and hold together large-scale social processes – such as Anderson's nationhood or Taylor's modernity – they take imaginaries mainly as ideational constructs. If the role of material constructs such as technology are discussed, it is only in passing. In Jasanoff's view, "Taylor's imaginaries do not have space for the material aspect of order" (Jasanoff, 2015a: 7). Considering this as a deficiency, Jasanoff aims to resolve it by providing a way to account for the interplay between the design of technologies and the social arrangements that inspire and sustain their production – in other words, how technology both embeds, and is embedded in, the social (Jasanoff, 2015a: 2–3). Jasanoff highlights the "instrumental and transformative" role that technological developments play in generating imaginaries of social order, that is, how society's self-reproduction relies on science and technology practices (Jasanoff, 2015a) and even more to the point, how imagination, material objects and technologies, as well

as social norms are joined together in practice (Jasanoff, 2015b) so that visions of technological progress contain also implicit understandings of what is desirable in our collective futures. By addressing the normative underpinnings of technoscientific developments, sociotechnical imaginaries give an explicit role for both the ideational aspects of societal transformations, as well as the material means of achieving those transformations.

Jasanoff defines sociotechnical imaginaries as “collectively held, institutionally stabilised, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology” (Jasanoff 2015a: 4). In Jasanoff’s view, these visions of how the world is and ought to be originate in dreams and ambitions of people, but are substantiated into objects and practices (Jasanoff, 2015b: 324). They can be articulated by nation-states (e.g., Felt, 2015), which was the focus in Jasanoff and Kim’s earlier iteration of the concept of socio-technical imaginaries (Jasanoff & Kim, 2009). Expanding from this notion, the context of the sociotechnical imaginary can also be a supranational organisation such as the EU (e.g., Mager, 2017). Smaller-scale organised groups may also orient their actions via collective imagination. These include, for example, corporations (e.g., Sadowski & Bendor, 2019; Smith, 2015) or social movements (e.g., Kim, 2015).

Yet another way to employ collective imagination in the context of technology is provided by the sociologist Patrice Flichy, who is interested in how subversive collective visions shape the realisation of large-scale technical objects (Flichy, 2007) – the case being the information network that became the internet. Flichy’s argument is that in the long term, such objects are articulated around a collective vision that is common to not only designers of technologies, but also the technology users. Flichy calls the collective imagination envisioning technological visions the *technical imaginaire*. Flichy builds on Ricœur’s (1986) above-mentioned schema of the interplay between utopia and ideology to describe the interplay between subversive elements of technology utopias, and legitimising and justifying elements of existing technological implementations. The utopian element of the initial technology vision means it aspires to alter the established order. It can still be broad enough to accommodate alternative technological implementations that can have contrasting features. The initial vision, then, does not represent an actual, realisable alternative to existing technical devices. Rather, there may be parallel realisable versions – in Flichy’s language, “project utopias” – that fit under the umbrella of the initial vision and can be experimented on. Once some of the experiments resonate widely enough among developers and users, they can catch hold and become to represent successful exemplars of the initial vision. The experiments are here successfully developed into boundary objects (Star & Griesemer, 1989) that represent a “compromise that can be used to associate multiple partners sufficiently loosely for everyone to benefit, yet

sufficiently rigidly for the device to function (Flichy, 2007: 10–11). This is where the ideological element of the exemplar now becomes visible. Its case-specific aspects are hidden, its success as a typical realisation of the initial vision is underlined, and the alternative realisations of the initial vision are cast aside. The exemplar may now be employed to encourage collective action of both technology developers and users. An initial utopian technology vision, then, gets worked through stages of experimentation on technical devices into material realisations that become to act similarly to an ideology; they justify the selection of specific solutions over others, and are able to mobilise developers and users.

## 2.3 Future imaginaries and economic dynamics

In addition to socio-technical imaginaries discussed above, another way to consider collective imagination in a specifically future-oriented way has been proposed by the economic sociologist Jens Beckert (2016). Here, the context is capitalism's success as an economic system, which Beckert attributes to economic actors temporally orienting their actions towards an open and uncertain future. This orientation is subject to the actors' capability to imagine the future and fill it with counterfactual economic imaginaries, which Beckert calls *fictional expectations*. Beckert discusses fictional expectations in an intersubjective sense, as a social phenomenon shaped by people's collective beliefs. Fictional expectations, then, refer to the images that actors collectively form of the states of the world in the future, and to the ways that actors expect their own actions, and the actions of others, to have an effect on those states of the world. These expectations are shaped by collective beliefs on how the economy functions, as well as by the cultural and institutional interpretive frames actors use to interpret economic events. In concrete terms, these collective beliefs and interpretive frames refer to, for example, economic theories and economic institutions, as well convictions such as reliance on technological progress as a solution to problems (Beckert, 2016: 13).

The effect that actions are expected to have in the imagined future are what motivate current action; in this way, actors behave “*as if* the future were going to develop in the way they assume it will, and *as if* an object had the qualities symbolically ascribed to it” (Beckert, 2016: 10; original emphasis). The future matters in the present, since actors' expectations of the future, and their expectations of how their current decisions will affect the future, will have consequences on their actions in the present. This leads to economic imaginaries having real-world consequences in the present, and through that, to the future state of the world, whether or not that effect is what was expected. Future imaginaries may be performative in nature, as has been discussed in the context of how economic theory shapes markets and their future outcomes (Callon, 1998; MacKenzie, 2006). Even if

the outcomes of current actions do not play out as imagined, the important thing is that fictional expectations motivate current real actions and therefore have a role in shaping the future.

For Beckert (2016: 11–12), fictional expectations have implications on the dynamics of capitalist economies due to their coordinating role and contingent nature. The coordinating role emerges in situations where actors have similar convictions regarding how the future will develop and the effects that their own, as well as others', actions and decisions have on it. One example of coordination are pressures institutionalised by the capitalist economic system, such as competition. Collective imagination motivates future orientation in two distinct senses. Actors seek self-advantage through new opportunities, but also experience the competitive threat posed by similarly advantage-seeking other actors. These pressures push actors to seek novelty through innovation and the increase of efficiency.

The contingent nature of fictional expectations, in turn, gives rise to a politics of collective imagination. Expectations can represent a departure from the current empirically observed reality; imagined futures make it possible to deviate from, for example, the present institutionalised economic practices or today's technologies. As expectations of the future affect present actions and through those actions, the future state of the world, it is possible to shape the future by shaping current expectations about it. Given that some future states of the world are more advantageous to a given actor, and given that future imaginaries of others have real-world consequences to the actor, actors have incentives to influence the collective imagination concerning the future to their own benefit. This gives rise to what Beckert calls the politics of expectations; future imaginaries are an entry point to economic power, and that is why they can become the object of present interest struggles (Beckert, 2016: 79–81).

## 2.4 Collective imagination as an analytical tool

The research aims of this thesis attend to the politics of collective imagination in the context of the data economy. The future of technology is created in the present through contested claims over technology's potential; therefore, examination of different imaginaries does not only reveal what different actors think about the future, but also tell us about what they want it to become (Wyatt, 2004). Imagined future states of the data economy motivate real, current decisions and become subject of current interest struggles (Beckert, 2016). If imaginaries are in this way able to affect current decisions and the future outcomes of the decisions, actors have incentives to attempt to affect the collective imagination in a way that suits their own interest. Collective imagination about the data economy, then, is a field of struggle



involving actors or groups whose economic and political interests, whether private or collective, are served by particular kinds of imaginaries.

My analytical approach is inspired by studies that use comparisons of imaginaries as an analytical tool. Comparisons are “perhaps the most indispensable method for studying sociotechnical imaginaries” as Jasanoff notes (2015a: 24). In this thesis, the comparisons take place between different data economy imaginaries serving different interests. Social imaginaries, as discussed by Taylor (2002; 2004), are focused on what holds things together rather than change, and therefore particularly attend to the ascendant imaginary instead of alternatives or conflicts (Mansell, 2012: 33). The social imaginary in Taylor’s sense may, however, be employed as a comparison point with which other imaginaries can be compared. Motivated by Jasanoff and Kim’s (Jasanoff, 2015a; Jasanoff & Kim, 2009) notion of multiple imaginaries and the emergent dominant ones, Robin Mansell’s (2012) analysis of dominant and alternative imaginaries of the information society, as well as the notion of counter-imaginaries by David Hess (2014), I will first establish the dominant imaginary of the data economy, consisting of taken-for-granted notions, images, and visions of those who are engaged in its development. Understood in this way, the dominant data economy imaginary consists of ideas about how the data economy functions as well as how it should function. Alternatives to this dominant imaginary act as conduits for “crystallising the dissatisfactions of the present into possibilities for other futures that people would sooner inhabit” (Jasanoff, 2015b: 329). An alternative imaginary emerging from discontent with the present is formed in response to the dominant imaginary, and works to alter the established order. I will therefore employ the dominant data economy imaginary for the purpose of assessing alternative imaginaries for the data economy through comparison.

There are two interrelated senses in which the politics of imagination is relevant for my analysis of different imaginaries through comparison, the first sense being the struggle between dominant and alternative imaginaries. Alternative imaginaries, here, being the ones developed in data activism. Taylor (2004: 30) notes that the constitution of a social imaginary takes place through the development of new practices, or modifications of old ones. The insight here is that the relationship between practices and social imaginaries works both ways. While practices are associated with social imaginaries, the transformation of practices can lead to the development of new imaginaries. Existing work on data activism by Stefan Baack (2015; 2018b) and Miren Gutiérrez and Stefania Milan (2019) has pointed out how the use and appropriation of data technologies in new ways by such “pioneering communities” (Hepp, 2016) can be the source of new data practices. The point of interest is that pioneering communities do not only develop new practices for themselves, but simultaneously work to expand the horizon of possibilities more generally. Broader transformations in practices, and hence social imaginaries, may

start from improvisation of new practices in smaller strata of the population (Taylor, 2004: 30); here, a pioneering community of data activists imagine how the use of technology can provoke change, exemplify new technology practices that others can adopt, and therefore “provide orientation for broader social discourses” (Hepp, 2016: 918). New practices emerging from data activism can shape new data economy imaginaries and instigate changes in the broader collective imagination.

The struggle between dominant and alternative imaginaries is not the only sense in which the politics of collective imagination is relevant in this analysis, the second sense being that different subversive visions for the data economy may embrace contested values and can work towards divergent, potentially mutually exclusive, alternative futures. This suggests that there can be *alternative* alternative imaginaries that may be in tension with one another, competing for power over a broader collective imagination not only against the dominant imaginary, but also between one another. While different alternatives can in the abstract sense work towards similar subversive visions, they may be underpinned by contested values and can be in tension with one another. In terms of technological development, these alternative imaginaries can end up producing contrasting or conflicting material realisations of the same or similar visions (Flichy, 2007). The success of these alternatives in the struggle over collective imagination, that is to say how well they succeed in expanding to broader social imaginaries, depends on how they fit together with existing norms and moral values, how they attach to things that generate economic or social value, how they resonate with collective identities, and how they tie in with existing social structures and material infrastructure (Jasanoff, 2015a; 2015b; Sadowski & Bendor, 2019).

In this thesis, the politics of collective imagination, then, concerns dominant and alternative imaginaries. Collective imagination will be employed as an analytical tool by making comparisons between imaginaries; particularly, comparisons between the dominant data economy imaginary and alternative data economy imaginaries developed in data activism. The dominant imaginary will be employed as the primary point of comparison, and comparisons against it will be made in order to reveal differences between the alternatives.

### 3 The Dominant Data Economy Imaginary

The aim of this chapter is to construct a view of the dominant imaginary about the data economy, which I employ as a comparison point for alternative data economy imaginaries. As discussed in Chapter 1, the data economy is in this thesis viewed as culturally, institutionally and politically embedded; the economy does not concern only market exchanges or even relations of production and exchange more generally, but also their broader cultural context, institutions, regulation and power relations. The data economy imaginary concerns collective beliefs and shared, possibly often implicit, understanding of how the data economy functions and what is normal and expected. Reflecting a broad view of the economy, these beliefs and understandings concern legitimate objects of economic interest and modes of operation, institutions governing production and exchanges relations, and the associated roles for economic actors.

The view of the dominant data economy is based on literature that I refer to as data studies, a stream of scholarship that “seeks to understand the new roles played by data in times of datafication” (Kennedy, 2018: 18). Data studies is not necessary a label that scholars would self-identify with, and strands of the literature I discuss might alternatively be referred to as critical data studies, surveillance studies, the political economy of data, or privacy research. Even if some of this literature concerns the politics of imagination in datafied times, the data studies literature does not usually discuss the data economy imaginary as such. Scholarship on data studies has, however, examined in detail the prevailing practices of producing and handling data, exploiting data as a resource for economic value creation, as well as the policies that govern these practises.

Data studies literature offers a view of the data economy’s dominant data arrangements which in turn reflect the dominant data economy imaginary: the taken-for-granted ways of producing or acquiring data, the means of making use of data to produce services and economic value, notions about monetising data and the associated market arrangements, the regulatory context that makes these practices possible, as well as the roles assigned for people as users, consumers and citizens in the contemporary digital environment. The examination of the dominant

arrangements can shed light to the dominant imaginary that underpins the data economy. For the purpose of my discussion, it offers a framing that allows the study of alternative data economy imaginaries.

### 3.1 The economic arguments for datafication

Contemporary organisations are culturally impelled to collect as much data as possible from all possible sources (Fourcade & Healy, 2017). The factors that drive this include beliefs and promises about possibilities to economically exploit data and the “radically new way of understanding and managing all aspects of human life” (Kitchin, 2014: 114) that they offer. In the industry parlance, data are presented as lucrative data capital that, if exploited correctly, can provide operational efficiency and economic advantages (e.g., Yousif, 2015) and firms can extract “big value from big data” (Gantz & Reinsel, 2011). Once “almost everywhere and nearly everything” have been datafied, “the potential uses of the information are basically limited only by one’s ingenuity” (Mayer-Schönberger & Cukier, 2013: 96). Because of this potential, data are “becoming a significant corporate asset, a vital economic input, and the foundation of new business models” (Mayer-Schönberger & Cukier, 2013: 16). All data are expected to have some insights hidden in them, it just takes the right application and the right analytical means to uncover them. It does not matter if all gathered data are not useful or valuable immediately; it makes sense to take hold of them, as by assumption they will ultimately turn out to be valuable (Fourcade & Healy, 2017).

These expectations of data’s value, combined with the statistical nature of firms’ analytic capabilities, lead to structural imperatives for increased scale and scope of data production, as well as to actors in the data economy valuing the quantity of data over their accuracy (Zuboff, 2015). More data, simply put, improves firms’ offerings; the more and more diverse data there are, the more and more diverse knowledge can be derived (Sadowski, 2019), leading to better service and more precise targeting (e.g., Mai, 2016). If more data leads to more valuable service, data are a competitive advantage, and a company benefits from data that others do not have (Srnicek, 2017). Accordingly, firms have an incentive to broaden the scope of their data collection to new people and new aspects of people’s lives, as well as an incentive to prevent others from gaining access to similar data. For example, the extraction of data takes place beyond the immediate boundaries of online services, an expansion of data extraction capabilities made possible by means of like buttons, tracking pixels and similar arrangements (Gerlitz & Helmond, 2013; Helmond, 2015). The need for new data extraction capabilities also at least partially explain the acquisition of up-and-coming competitors by data giants, as well as their forays into new fields such as payment systems or health applications.

When data are considered as a form of capital, they may be turned into value in various ways. Sadowski (2019) identifies six such ways, presented here as one taxonomy of deriving value from data. First, and perhaps foremost when it comes to personal data, data are used to profile and target people. As mentioned in Chapter 1, a considerable share of revenue of some of the most successful internet firms come from personalised advertisements. Their business models are driven by the notion that knowing more about people translates into more value through more nuanced profiling and better targeting. Businesses also translate knowing more about people into value and profit by scoring people in industries such as credit and insurance (Fourcade & Healy, 2017) and by means of targeting and price discrimination in the retail industry (Turow et al., 2015b). The data broker industry specialises on creating dossiers of data about people and offering them to different kinds of businesses for the purposes of knowing more and better (Crain, 2018).

The second mean of deriving value is using data to optimise systems (Sadowski, 2019) for more efficient processes and employment of capacities, for eliminating waste or downtime, and for the improvement of productivity. For personal data, examples include making workers more efficient in warehouses, and optimising customer service processes or the running of public services. Third, value is derived from data by using them to manage and control things. Here, the ability to exercise power over something is enhanced by amassing data on it. To manage health, the data in question might concern exercise, diet or bodily functions; to manage traffic in a city, the data may concern the environment or the movement of people and things; to manage and control clicks on websites, the data may concern browsing patterns collected via tracking cookies. Fourth, data are used to model probabilities of future events, such as risks of illness for the purpose of preventive healthcare, or the likelihood of criminal activity by individuals or groups for the purpose of predictive policing (Andrejevic, 2017). Fifth, data are used to build things, such as services or digital systems (Sadowski, 2019). A service like Uber would not work without real-time data about users and drivers. Similarly, smart homes or smart cities depend on personal data, among other kinds of data. Finally, data about the use of assets such as buildings or machines is used to prevent or slow down the depreciation of their value over time. These means of deriving value from data are interlinked. For example, the modelling of probabilities of future events may be related to knowing more about people, the management of things, or the optimising of systems. What they have in common is that they all represent what Cinnamon (2017) refers to as “competitive value”; the value lies in the possibility to leverage data for competitive advantage over others.

When data are employed by organisations as a form of capital, they are naturalised as a universal substance or a generic raw material that is there for the taking, and that can be used or turned into value (Couldry & Yu, 2018; Pendergrast,

2019; Puschmann & Burgess, 2014; Sadowski, 2019). The beliefs and assumptions that underpin this use of data to understand human life are summarised by van Dijck (2014) as constituting dataism, or the ideological component of datafication. A central belief related to dataism is that datafication is a legitimate means to access, understand and monitor behaviour. To be viewed as a raw material that can be turned into meaningful predictions about people's actions and behaviour, data are considered an objective representation of the world, and platforms and services on which data are collected are channels through which such data can be acquired. Consequently, data are believed to represent human behaviour in a manner that can be meaningfully mapped to people, and represent social life in a manner that social identities, relationships and practices can be reduced to data and, through data, processed as abstractions (Hintz et al., 2019: 46–47). Data collecting and processing technologies are therefore viewed as neutral conduits that can be used to uncover truths that exist regardless of data about them. Supported by these beliefs of objectivity and neutrality, the promises of value and competitive advantage have turned datafication into business-as-usual mode of operation in the data economy.

From political-economic and philosophical perspectives, however, these assumptions about datafication and its benefits are set up on highly problematic grounds. Data are not simply “raw” or “objective,” but rather always a result of various different decisions and assumptions that go into their production. These decisions – and, by extension, who makes them – can be consequential to the outcomes of data use. Data does not simply exist out there waiting to be harvested, rather it is produced by technology and people using and creating that technology (Gitelman & Jackson, 2013). When people's behaviour and social life are represented by data, complex things are mapped to simpler ones, dimensions of the situation are truncated and projected, and choices are necessarily made about how to carry this out. When data about people are employed and analysed to generate representations of social identities, data technologies are “embedded and integrated within a social system whose logic, rules, and explicit functioning work to determine the new conditions of possibilities of users' lives” (Cheney-Lippold, 2011: 167). Rather than uncovering existing truths or knowledge, technologies used to produce and process data contribute to the creation of certain kinds of regimes of knowledge (Monahan, 2008).

Irrespective of these problematic aspects, the arguments about the benefits that data-based new means of knowing offer firms constitute a powerful discursive regime (Kitchin, 2014) that provides the rationale for adopting data technologies and legitimises their development, making its message commonsensical and persuading actors to its logic. As Kitchin (2014) points out, countering the narrative about datafication's economic benefits is difficult. If datafication promises lucrative business models and competitive advantage through optimisation, efficiency,

control, foresight, new innovative services and durable assets, who would want to be left out? Rejecting data's value means having less of these things, in other words, missing out on innovative services, being less efficient and giving competitive advantage to others.

### 3.2 Datafication and informational capitalism

The collection of data about people has always been important to the practices of governance and control, but the questions around data and power are becoming more central due to the advancement of datafication (Hintz et al., 2019). While the accumulation of data was previously by and large the domain of the state, we are now witnessing an expansion of the role of private sector in this domain (Andrejevic & Gates, 2014; Flyverbom et al., 2017). Accordingly, a growing literature examines data-based business practices from the point of view of political economy of datafication, focusing on the power relations shaping the production and distribution of data, and on the economic logic of data-based knowledge production. So far, this literature has not converged into agreed-on concepts and terminology, and authors discuss contemporary capitalism in the digital environment with a focus on, for example, data as a form of capital, surveillance as a mode of producing data, and platforms as the apparatus of data production. The use of different concepts and approaches reflects parallel, potentially complementary and partially overlapping emphases, some of which do not acknowledge one another. What is in common with these approaches is that they consider datafication as a mechanism driving a new stage of capitalism, a specific form of informational capitalism (Castells, 1996) that is quickly becoming a significant underpinning of the contemporary society. Taken together, the literature analysing business data practices and data arrangements of the private sector is a description of the contemporary data economy imaginary.

One of the commonly employed concepts in this literature is *data capitalism*. The concept broadly stresses the economic role of datafication, the commodification of data (Hintz et al., 2019) and data as a form of capital (Sadowski, 2019). For example, West (2019) examines data capitalism as a result of construing data as legitimate object of economic interest, tracking data capitalism's development as a consequence of turning from an e-commerce model, premised on the sale of goods online, to the sale of advertisements based on behavioural profiles tied to personal data. Data capitalism's economic logic "places primacy on the power of networks by creating value out of the digital traces produced within them" and "appeals to community and consumer power to mask the digital labour it relies on [and] calls into question the conflict between our needs for privacy and desires for community" (West, 2019: 21).

*Surveillance capitalism* as discussed by Shoshana Zuboff first in an influential article (2015) and later in a polemical book (2019), has become another key concept. Zuboff outlines the features of “a new economic logic based on fortune-telling and selling” (Zuboff, 2019: 88) that hinges on extracting data about individuals “for the purpose of knowing, controlling, and modifying behaviour to produce new varieties of commodification, monetisation, and control” (Zuboff, 2015: 85). Some of the data that firms extract about people are employed to improve service provided to people. Data in excess of this is, in Zuboff’s metaphor-rich terminology, behavioural surplus. For Zuboff (2019), “normal” and acceptable informational capitalism based on personal data turns into more dubious surveillance capitalism when behavioural surplus is employed to know more about people’s behaviour now and in the future, for the purpose of using that knowledge as a means to serve not people themselves, but others’ ends. The focus in Zuboff’s writing is on this specific, questionable mode of data exploitation, the economic incentives that drive surveillance capitalist firms to increase behavioural surplus, and on the social relations that the economic logic implies. Zuboff’s analysis highlights the mechanisms of data extraction, behaviour prediction, commodification of behavioural modification, and control, which she identifies as enabling the production of new markets for data-based products and services but hamper agency and self-determination by “exiling” people from their own behaviour.

Other, less widespread concepts have also been employed to analyse contemporary informational capitalism. Scholarship on *drone capitalism* (Andrejevic, 2016; Richardson, 2018) focuses on the cascading logic of automation, describing how automatic data collection leads to the necessity of automating analysis, then decision-making, and ultimately action. Each step of automation seems to naturally and irrefutably lead to the next step due to scale increases, making it difficult to contest the logic. Some of the more critical accounts examine data relations through the lens of *data colonialism* (Couldry & Mejias, 2019; Thatcher et al., 2016), likening informational capitalism’s appropriation of data and the normalisation of their economic exploitation to “the predatory extractive practices of historical colonialism” (Couldry & Mejias, 2019: 336).

Datafication as a basis for value creation and monetisation is perhaps nowhere as obvious as in digital platforms. Authors discussing *platform capitalism* (e.g., Srnicek, 2017) focus on digital platforms and their role in contemporary informational capitalism in arranging interactions so that data can be efficiently produced. Theoretically, two perspectives to platforms can be identified: from an engineering perspective, they are modular technological architectures; and from an economic perspective, they act as intermediaries of multi-sided markets (Gawer, 2014). In a discursive sense, a strict understanding of platforms as either an engineering or an economic phenomenon has long since been relaxed to favour a



more everyday understanding as online services of various intermediaries (Gillespie, 2010). Platforms make two or more groups meet and allow them to interact. To name some possibilities, users, customers, advertisers, application developers, content providers or service suppliers (Srnicsek, 2017). Consumer-facing intermediaries include Google search, Facebook, Twitter, Uber or AirBnB. However, also Amazon Web Services or “industrial internet” services that embed computational and communicational capacities within manufacturing processes can be understood as digital platforms. Being situated in the locus of interaction between different actors or groups, platforms can also position themselves as the substrate where the interactions occur. It is, therefore, possible to datafy these interactions, and the platform can serve as the “extractive apparatus of data” (Srnicsek, 2017: 48).

Two economic characteristics are essential to the efficiency of platforms as the data extractive apparatus: network effects and cross-subsidisation. Services affected by network effects get more valuable to everyone as they get more users. In digital platforms, network effects can work by two mechanisms. The service can simply become more enticing as more users are involved; a social networking site becomes more useful the more participants it has, and similarly AirBnB improves along with more offerings to choose from. Network effects may also apply to data. The more users a service has, the more data it can collect, and the more raw material there is for personalisation or recommendation systems. If improved personalisation and recommendations lead to more users choosing the service, more data can again be collected which leads to even more improvements, still more users and so on (Rieder & Sire, 2013; Srnicsek, 2017). Cross-subsidisation, in turn, refers to how platform companies subsidise losses incurred in one market in order to stimulate sales in other, profit-turning market (Rochet & Tirole, 2003). The purpose is to better leverage network effects. If the increase of users on one market makes services offered on the other market more valuable, there can be an incentive to cross-subsidise. In practice, consumer-facing online services often offer services for free to people, and expect profits from customers in markets where targeting service is sold to businesses.

The above accounts of contemporary informational capitalism describe how firms encompass datafication on the mass scale within an economic logic. One of their main outcomes is that the emergent new form of informational capitalism results in a new asymmetric distribution of power. As Sarah Myers West summarises, while “communication and information are historically a key source of power (Castells, 2007), data capitalism results in a distribution of power that is asymmetrical and weighted toward the actors who have access and the capability to make sense of data” (West, 2019: 23). In contemporary informational capitalism, knowledge that is produced based on personal data is shaped to serve the interests of those actors who have the power in the sense of having the means to make use of data. Firms exploit this position of power to construct and control the markets in

which the exchanges of data, value and money take place, and in the context of competition among firms, do this in a way that advances their own economic interests. Consequently, markets which determine who benefits from personal data are by design arranged such that users do not participate in them. This does not mean that people would not benefit from the use of their data; we obviously do, for example, in the form of highly useful digital services. Rather, the point is that the ways in which people can benefit are determined in the context of markets and exchanges that are shaped for firms' profit-maximisation, and designed so that users do not participate in them. In short, the answers to the questions about how, by whom, and for what purposes knowledge is produced are already shaped by the imaginary of how the data economy functions.

### 3.3 The social relations of data extraction

Another commonality of the analyses of data economy's economic logic lies in how they frame the production of data as data extraction (e.g., Couldry & Mejias, 2019; Sadowski, 2019; Srnicek, 2017; Zuboff, 2015; 2019). Terms such as "collection" or "gathering" conjure the image that data are already out there available for the taking (Sadowski, 2019: 6), and references to data traces or "data breadcrumbs" (Edge, 2012) suggest that people only accidentally leave data behind as they go about their daily business. Extraction, in contrast, brings the focus to the ways that people are targeted with purposeful processes that produce data. These processes are asymmetric in the sense that they tend to go unobserved by those targeted with them, and to produce data for purposes unknown to them (Andrejevic, 2014).

Surveillance studies has a history of studying processes similar to data extraction, and this literature provides insight into the nature of data extractive processes. Surveillance, in a conventional sense, is defined as systematic, routine, and focused attention to personal details for a given purpose (Lyon, 2007; 2014). The purpose in question may include, for example, management, control, influence, or protection. When information technology is employed to automate surveillance, some authors (e.g., Raley, 2013; van Dijck, 2014) employ the concept of *dataveillance*, referring to surveillance by means of personal data systems (Clarke, 1988), potentially on the mass scale to monitor large groups of people (Clarke, 2003). *Dataveillance* tremendously enhances the efficiency of surveillance. As Raley (2013: 124) notes, it is not only that there is a difference in degree when *dataveillance* makes it more efficient and simultaneously less visible to systematically focus attention on personal details. There is also a qualitative difference; surveillance by means of *dataveillance* is not only descriptive, but also predictive and prescriptive (Raley, 2013) – that is, *dataveillance* makes it possible to not only monitor current and past behaviour, but also to form predictions about future behaviour, intentions,

and characteristics. In addition, predictions of future behaviour enable intervening in behaviour before the fact (Lyon, 2014) and to even modify behaviour, for example by constructing personalised choice environments that do not necessarily enforce or restrict choices, but rather nudge towards choices preferred by the designers of the choice environment (Yeung, 2016).

Another qualitative difference to traditional surveillance is that dataveillance is not selective (Andrejevic & Gates, 2014; Lyon, 2014; van Dijck, 2014). Where surveillance has conventionally relied on targeting specific persons of interest, in dataveillance there is no difference between targets and non-targets of data collection (Andrejevic & Gates, 2014), since predictive processes require knowledge about both groups in order to reveal differences between them. Likewise, where surveillance presumes monitoring for specific purposes, dataveillance can be automatic and continuous, performed for unstated purposes that only become known later on (van Dijck, 2014). Initial non-selective dataveillance allows the selection of targets for intervention later on.

Mark Andrejevic and Kelly Gates (2014) have identified three structural features of data extraction that have implications for social relations in the digital environment under the dominant data economy arrangements. The first of these is opacity, which can be thought to stem from data being inferentially fertile (Manson & O'Neill, 2007: 104). This is to say that data do not only contain specific pieces of information in a straightforward manner; instead, different correlations in and across different datasets may be uncovered by inference and data mining. These correlations may be un-intuitable and it can be impossible to find a common-sense explanation for them. However, it is, or at least it is imagined to be, enough to simply make use of their predictive power. Because of this opacity, knowledge derived from data is epistemologically oriented to serve those who produce them, rather than the people whom they target (Andrejevic & Gates, 2014). The second structural feature, speculativeness, refers to the potential value of the uses that data may be put in the future. Speculativeness stems from structural opacity; it is not possible to know the actual and potential uses of data *ex ante*, as new, valuable correlations may be found only when data are combined with other data. Therefore, justification to data extraction may only be presented *ex post* (Andrejevic & Gates, 2014). This leads to incentives to consider all data as signals to be analysed (Mayer-Schönberger & Cukier, 2013); as much data as possible should be recorded before even determining the full range of their potential uses (Lyon, 2014: 4). The third structural feature, asymmetry, stems from how dataveillance is reliant on infrastructural capacities to extract data and produce predictions (Andrejevic & Gates, 2014). It relies on the availability of data, as well as specialised means of production relying on proprietary expertise and capabilities (Zuboff, 2015) such as data management, automated data processing and computing power. Data-based knowledge is available only to those

who are capable of tapping into these resources. This asymmetry has been elsewhere characterised by Andrejevic as the “big data divide” (Andrejevic, 2014); data extraction does not only signify the separation between individuals and their data, but also a separation in the sense of the ability to analyse and make use of the data.

What these structural features of data extraction imply for social relations in the data economy is that they institutionalise the lack of reciprocities between those who extract data, and those who are the targets of extraction. Even though personal data signal personal and potentially intimate details about people, data extraction is a one-way process occurring in the absence of dialogue between people and those who extract data (Zuboff, 2015). Extracted data are assembled into representations of people that are constructed for the purpose of intervention (Haggerty & Ericson, 2000); however, the intention of creating this representation influences how it is constructed, what data it includes or left out, and the kinds of intervention it enables (Dalton et al., 2016). Even if users are aware of data extraction and the consequent production of knowledge, these data and knowledge are not oriented to serve their interests.

### 3.4 Formal institutions regulating data extraction

Of the formal institutions providing collective rules of interaction between people and organisations doing data extraction, privacy-related regulation is perhaps the most obvious one. While there are differences between jurisdictions, an important regulatory approach both in the EU as well as the US is to protect people’s informational privacy based on the principle of informational self-determination, under which people are charged with making choices about personal data (e.g., Coll, 2014; Solove, 2013). In practice, this means that data extraction requires providing people with information about it and asking their permission for it, a practice that we often come across in our daily lives, whether on- or offline. Under this regulatory approach, the right to informational privacy is a right to decide, and the exercise of this right means choosing one’s position in the spectrum between secrecy and transparency (Zuboff, 2015). The assumption is that “data subjects make conscious, rational and autonomous choices about the processing of their personal data” (Schermer et al., 2014: 171). If the benefit expected from revealing data outweighs the associated harm, an individual performing rational cost-benefit calculations would choose to reveal the data (Solove, 2013).

Consequently, when data are extracted in accordance with privacy regulation based on informational self-determination, this would be a voluntary process on the part of the data subject. That companies provide their users with services and simultaneously extract data about them would, then, indicate that people have grown accustomed to the trade-off between extraction of data and free services (van Dijck,

2014). This would mean that people are comfortable with the exchange they are presented with, and consider the benefits they receive to outweigh the costs. Firms emphasise exactly this position; users always have the choice not to use their services, and that what we are witnessing is indeed people's rational acceptance of trade-offs based on cost-benefit calculations (Schermer et al., 2014; Turow et al., 2015a). However, empirical research based on surveys, experiments, focus groups and interviews has established that privacy emerges as citizens' primary concern in the digital age, with people complaining that their rights and the ability to control their personal data are limited (e.g., Hargittai & Marwick, 2016; Hoffman et al., 2016; Kennedy et al., 2015; Marwick & Hargittai, 2018; Selwyn & Pangrazio, 2018). Despite this, people often reveal private information for even small rewards and engage in dataveillance practices as voluntary participants, "prosumers" (Ritzer & Jurgenson, 2010) of sorts. To put it differently, although when asked, people claim to desire for and care about informational privacy, they nevertheless behave in ways that contradicts those claims.

The inconsistency between expressed attitudes and observed behaviours is often called the privacy paradox (Norberg et al., 2007), and different explanations for the paradox have been provided (Draper & Turow, 2019; Kokolakis, 2017). One explanation was already described above: we are observing informed people who engage in privacy calculus, recognise privacy harms, and in the cases where the apparent paradox occurs, simply judge the benefits accruing from data collection to be larger than the associated harms (see Draper, 2017; Hoofnagle & Urban, 2014). Another explanation focuses on the bounds of rational decision-making that delimit privacy calculus (see Acquisti et al., 2015): people attempting to perform privacy calculus are affected by heuristics, emotions or misperceptions of costs and benefits of data disclosure, which cause them to misjudge the situation. A third type of explanation stresses that people are uninformed of the ways personal data are collected and used, and therefore simply unaware that they engage in behaviour that contradicts their stated intentions and preferences (see Barnes, 2006; Dommeyer & Gross, 2003; Park, 2013). An imaginative explanation based on quantum indeterminacy has also been offered: privacy preferences, like the observables in quantum physics, remain indeterminate until they are observed when the actual privacy decision is made (Flender & Müller, 2012).

Privacy research has also highlighted various ways in which privacy calculus is not easily performed despite best of intentions. People are expected to provide consent for data extraction in conditions characterised by lack of transparency, context-dependent and malleable attitudes towards privacy, and oftentimes a lack of real choice over whether or not to reveal data (e.g., Acquisti et al., 2015; Schermer et al., 2014; Solove, 2013). Cost-benefit analysis on the exchange of data for services requires comparing two abstract things, the valuation of which is complicated to

begin with. To further complicate things, even if people had control over data they do voluntarily provide, they would not have control of the data that is produced on the basis of other data (Mai, 2016). Further, the terms of data extraction are largely imposed on the users (Degli Esposti, 2014), and the choice environment can be engineered to serve specific interests and to make meaningful consideration of privacy decisions difficult (Crain, 2018; Draper & Turow, 2019; Monahan, 2016). The framework based on informational self-determination is also fundamentally challenged because it focuses on individual choice and therefore fails to acknowledge collective dimensions of privacy (Baruh & Popescu, 2017). Due to these reasons, many have argued, it is challenging or impossible for people to make meaningful decisions on personal data, regardless of intentions to do so (see Article I).

Sami Coll (2014) argues that a concern over loss of privacy, combined with a continuing understanding of privacy in terms of self-determination, leads to the main project becoming the education of people to protect themselves. People become to be seen as the first line of defence in need to be enlisted to guard their own privacy (e.g., Norberg et al., 2007: 120). A combination of the structural incapability for meaningful self-determination with a regulatory approach that nevertheless relies on self-determination can be viewed as the culprit of efficient dataveillance practices; people are nominally in control of their data and privacy, but some of that control is in practice transferred to those doing dataveillance. In other words, if the structural features of dataveillance lead to privacy self-determination being possible to only a limited extent, “the notion of privacy and the surveillance of data act as the ‘partners-in-crime’ of the current growing digital economy” (Coll, 2014: 1253). Companies and organisations ask users to provide broad permissions for extracting and using data, and therefore simultaneously accumulate rights to make decisions on data use (Zuboff, 2015). If privacy is a decision right, dominant data arrangements redistribute these rights, concentrating them from people to companies.

The development of dominant data arrangements has taken place ahead of meaningful new regulatory developments that could curb them. In light of how privacy protection acts as a force of co-opting individual control for the purposes of enacting efficient forms of data extraction, their development is not necessarily delimited, but rather made possible, by existing privacy regulations. In this sense, privacy regulation acts as an institution making the prevailing data arrangements possible. Regulation provides roles for people and the entities producing data, and rules for interaction between them, and these roles and rules work to make prevailing data extraction practices possible.

### 3.5 Resignation towards dataveillance

People are the ultimate sources of personal data, and could in principle demand alternative ways of organising data relations in the digital environment. As discussed above, some explanations for why people continue to voluntarily participate in and allow data extraction have been uncovered in research on the privacy paradox. An alternative explanation for people's apparently voluntary participation can be framed in terms of politics of collective imagination about dataveillance. I will here draw from two recently proposed sociological concepts, surveillance realism (Dencik, 2018) and digital resignation (Draper & Turow, 2019).

By surveillance realism, Lina Dencik (2018; also see Dencik & Cable, 2017; Hintz et al., 2019) refers to the contemporary social condition of the digital environment: the normalisation and common-sense nature of ubiquitous dataveillance, and the experienced inurement to prevalent dataveillance practices, whether by state or commercial actors. Dencik builds on Mark Fisher's (2009) notion of capitalist realism, a framework for viewing capitalism's effects on cultural production, politics, economics and thought. Capitalist realism refers to the sense in which capitalism is not just viewed as the only viable political and economic system, but also normalised to an extent that it seems to have become impossible to imagine alternatives to it. Dencik makes an analogous argument about surveillance in the digital domain. As an example of developments leading to the condition of surveillance realism, Dencik presents post-Snowden protests against state dataveillance, which did not result in widespread responses from the broader public due to a securitisation discourse that worked to justify and normalise the dataveillance practices as necessary responses to imminent security threats. Through such developments, and despite simultaneous widespread unease of dataveillance practices, infrastructures and systems, people have internalised their necessity (Dencik & Cable, 2017). Dataveillance infrastructures and practices have become normalised to the extent that their existence is the "pervasive atmosphere" (Fischer, 2009) that limits thought and hampers possibilities of imagining alternative ways for organising relationships in the digital environment.

In the condition of surveillance realism, people are resigned to the present state of affairs, and simultaneously alternatives do not seem plausible. To explain just this resignation on a detailed level, Nora Draper and Joseph Turow (2019) draw from a set of empirical findings on people's attitudes towards data collection and privacy. These include a pervasive feeling that corporate data arrangements are unfair but a simultaneous sense of resignation over the matter (Turow et al., 2015a); people's feeling of powerlessness to avoid privacy violations (Hargittai & Marwick, 2016); a reduced sense of control induced by what is seen as compulsory disclosure of data (Marwick & Hargittai, 2018); and the recognition of privacy risks and simultaneous uncertainty, mistrust and lack of power that render privacy protection behaviour

subjectively futile (Hoffman et al., 2016). These findings indicate that individuals engage in privacy protection or the opposition of dataveillance, but feel that those efforts are generally unsuccessful (Selwyn & Pangrazio, 2018). People desire to have control towards information about them, but feel unable to exercise it in a meaningful way. The inurement to dataveillance and the inability to meaningfully contest it are, according to Draper and Turow, key elements of the contemporary social condition in the digital environment. This social condition has not developed arbitrarily; instead, it is cultivated by specific, widespread corporate behaviours and communicative strategies that work to obfuscate the situation and convey a sense of normalcy around dataveillance practices (Draper & Turow, 2019). Examples of this cultivation include technological systems that mislead users by presenting them with the illusion of control (Monahan, 2016); transparency initiatives that suggest user empowerment to control data but in practice give little insight to the firm's actual practices (Crain, 2016); and privacy policies that require extensive time to be familiarised with (McDonald & Cranor, 2008) and are difficult to understand due to jargon that limits comprehension and discourages careful reading (e.g., Pollach, 2005).

Both Dencik as well as Draper and Turow stress that the inurement does not imply that people would be passive or apathetic towards dataveillance. It is, rather, that their attitudes towards dataveillance reflect, and are negotiated in the context of, the way dataveillance systems are integrated in the modern society. These attitudes are nurtured by data-using companies by means of specific data arrangements in order to ensure continued possibilities of data extraction. The outcome is that people's desire for more control does not lead to motivation to work towards change; people consider meaningful engagement with data arrangements and their politics too time-consuming, demanding, and ultimately futile.

### 3.6 The data economy imaginary

As Charles Taylor (2004: 25) suggests, practices and imaginaries should be understood to go hand in hand; here, the data arrangements that dominate our digital habitat go hand in hand with the dominant imaginary underlying the spread and application of data technologies. This is to say that the data arrangements discussed in this chapter should be simultaneously understood as the tacit understanding of how the data economy functions, what is normal and expected, and how things usually go between data economy's actors. This understanding, in turn, is embodied in data economy actors' taken-for-granted processes, procedures and ways of doing things. These data arrangements have become "a key underpinning of contemporary society, necessary to the successful operation of the economy [...] Data collection practices are not merely normalised as inevitable component of information



infrastructures, but also justified through discourses circulating within the spheres of policy-making, media coverage and everyday interactions” (Hintz et al., 2019: 10).

Based on this discussion, I highlight two interconnected aspects of the dominant data economy imaginary. The first aspect concerns the drive for the extraction of more data. Through dominant data arrangements, datafication is tied in with firms’ notions on how data and datafication are normally exploited economically. This is to say that certain cultural routines concern data arrangements, certain ways of producing data and new knowledge based on data are viewed as legitimate, and there are certain normal, taken-for-granted ways of using data in the production and monetisation of economic value. Data are viewed as a resource there for the taking, a form of capital that can be legitimately exploited in the creation of economic value. Data are associated with more efficient ways of conducting existing business, with a potential for highly successful, innovative and disruptive new forms of conducting business, with increased control of assets, customers, employees and processes, as well as with foresight into the future. These promises of data’s value are competitive in nature, in the sense that they contain promises of competitive edge against others. Conversely, neglecting data’s opportunities would mean less innovation, less efficiency, less control, and less foresight. Once this imperative is accepted, market competition orients companies to hone their systems – digital platforms, market relations, and policies that govern them – such that as much data as possible can be extracted, in ways that are deemed acceptable in view of the established data arrangements.

The second aspect concerns the part that people play in the data economy as citizens, consumers and users. What has become the taken-for-granted shape of the data economy affects not only how companies operate, but also how digital technologies permeate people’s lives. The role for people is not that of a consumer making choices, nor an agentic citizen; rather, people are the sources of data and targets of data extraction. How data are used is decided on the basis of how they best serve those who have extracted them; these decisions are normally made in the context of markets facing paying customers, largely other businesses. These arrangements are made possible by the prevailing formal institution of privacy regulation. In the dominant data economy imaginary, the taken-for-granted response to the question of who decides what data are collected, what is learned based on it, who does the learning, and who decides who benefits, is “not people themselves”. The structural features of dataveillance – opacity, speculativeness, and asymmetries – tend to ensure maintaining “one-way mirror social relations” (Zuboff, 2019: 81) between firms and users. The operations and markets in the data industry are incompatible with the idea of empowering individual people through increased transparency and control (Crain, 2018), and the forms of knowledge generated by dataveillance “push against any attempt to delimit either the collection of data or the

purposes to which that data are turned” (Andrejevic & Gates, 2014: 192). Dataveillance permeates the digital domain to the extent that it is difficult to imagine alternatives to it, and companies cultivate the condition of digital resignation to maintain the present, taken-for-granted shape of things in the data economy (Dencik, 2018; Draper & Turow, 2019).

The dominant data economy imaginary is grounded on beliefs about datafication and about data arrangements that orient them to serve primarily companies’ interests, and this imaginary works to enable the continuation and expansion of data extraction. The two aspects highlighted above point out how the data economy imaginary is not just the imaginary of those commercial actors that utilise data for their own purposes. Rather, through how data economy’s arrangements permeate people’s lives, the dominant data economy imaginary also becomes to affect how people view datafication and how they expect things to normally work in the data economy.

## 4 Data Activism

Langdon Winner notes in *The Whale and the Reactor* that our thinking is driven by an “almost religious conviction that a widespread adoption of computers and communications systems along with easy access to electronic information will automatically produce a better world for human living.” (Winner, 1986: 105). Indeed, the use of technologies that convert aspects of social life into quantifiable data has been largely accompanied by a rhetoric stressing their political and societal benefits (e.g., van Dijck & Nieborg, 2009; West, 2019). Similarly, much writing related to citizenship in the digital domain has been focused on citizens’ agency and empowerment that can be derived from the use of data-enabled digital tools (see Hintz et al., 2019: 21). However, like technologies generally (Feenberg, 1999: 7), data technologies are ambivalent in the sense that they are available for alternative developments with different social consequences. This is to say that with the widespread adoption of data technologies, we might or might not end up with a better world for human living. Citing Andrew Feenberg’s philosophy of technology, Helen Kennedy (2018) argues that what fundamentally matters for these alternative developments is the subject position of people in relation to data technologies: whether we are dominant or subordinate to technological systems. When looking at technological systems, the issue is not with just what can be derived from the use of data-enabled tools, but also the subject position citizens have in the data arrangements in which those tools are embedded.

The development of data technologies has largely taken place under the radar, outrunning our understanding of what is at stake, meaningful ethical scrutiny, or functioning regulation that could come along with them (Zuboff, 2015). Recently, however, straightforward notions of citizens simply deriving agency and empowerment from digital tools have arguably been complicated by a gradual collective waking up to the political economy of dataveillance. Even if data technologies have thus far developed in a yet un-normed space, “the realm of human activity outside the state and the market [...] is now slowly but steadily catching up and turning ‘big data’ to its own ends” (Milan & van der Velden, 2016: 59).

As one outcome, data activism has emerged as a distinct category of civic engagement with data technologies. Data activism, broadly understood, refers to

civic engagement and political action that in some sense emerge in response to the uneven distribution of data access and data use capabilities in datafied times (e.g., Baack, 2015; Beraldo & Milan, 2019; Gutiérrez, 2018a; Milan & Gutiérrez, 2015; Milan & van der Velden, 2016; Schrock, 2016). Data activism has emerged alongside increasing citizen awareness of different kinds of potentials that data have: for social change, for surveillance, for economic exploitation or benefit, for control of others or the self. As will be discussed below, data activism may aim to interfere with, hijack, contest or re-appropriate (Beraldo & Milan, 2019) existing data arrangements. The aim can be preventing the use of data for the benefit of others, or making use of data to improve the citizens' own condition. In other words, data are brought forward as the object of intentional action. Data activism, as Milan and Gutiérrez put it, "brings back into the data collection machine the fundamental elements of agency and politics" (Milan & Gutiérrez, 2015: 130); it is about citizen agency in datafied times.

If technologies are ambivalent in the sense that they are available for different social consequences, the dominant data economy imaginary outlined in the previous chapter is but one possibility for how data technologies affect and permeate our lives. New data arrangements developed in data activism can amount to reconsideration of relations between technologies, people, firms, and states in the production of data, knowledge and value. In other words, they can act as underpinnings for an alternative data economy.

This chapter discusses the continuously expanding scholarship on data activism, with a purpose of providing an overview of the ways in which data activism relates to the dominant data arrangements in the data economy. In addition, the aim is to situate the MyData initiative, described in Chapter 1 and empirically examined in Articles III and IV, in the broader context of data activism.

## 4.1 Civic engagement and the digital environment

Whereas data activism as such is new phenomenon due to its connection to novel data technologies, the social forces driving it are not new (Milan & Gutiérrez, 2015). Data activism may be viewed as a recent variety of advocacy, grassroots political action, and appropriation of information and communication technologies that are both motivated and made possible by digital technologies. These forms of civic engagement are connected with people's rights, capabilities and roles as users, consumers and citizens – that is to say, they broadly concern digital citizenship, or citizen agency in the digital environment (Hintz et al., 2017; 2019). These forms of advocacy and citizen engagement indicate different ways of reacting or responding to the closing off and monopolising of knowledge production and value creation in digital environments.

Civil society's engagement with, and in, the digital environment can take many forms. Some entail collective organisation of advocacy; nowadays the use of digital media naturally forms the basis for many forms of civic engagement, such as community building and organising work for both digital-specific campaigning and more general political causes (see Kaun & Uldam, 2017). Digital-specific political campaigns include, for example, digital rights activism that employs internet to defend values that are seen to lie at the heart of internet in a recursive manner (Breindl, 2013); the freedom of information movement that advocates communication freedoms, publicness of information, and state and corporate transparency, and the liberalisation and homogenisation of regulations across jurisdictions (Beyer, 2014); and the advocacy for rights of consumers to participate in the production of cultural content, promoting recognition of users' agency in creating products that shape mass culture, and their legitimate claims over the content they create (Postigo, 2012).

Other forms of citizen engagement entail the development of technologies in order to develop alternative forms of material culture. One example is the free software movement investigated by Christopher Kelty (2008), already mentioned in Chapter 2. The free software activism is based on a particular relationship between technology, governance and the internet, so that enacting societal change is combined with ideas about the correct role of technology in achieving that change. Sometimes promoting alternative forms of technology is directly combined with forming connections or collaborations with the private sector. In such cases, social change is sought through technologies that are developed and produced by commercial firms, the "open source" variant of the free software movement being one example (Hess, 2005).

Highlighting the importance of data and information in contemporary politics, hackers and hacktivists act to counteract the power of governments to shape the internet and limit freedoms (Milan & Gutiérrez, 2015: 122). Civic hackers, for example, building on the ideas of freedom of information, deploy infrastructure and tools as a mode of "requesting, digesting contributing, modelling and contesting data" (Schrock, 2016: 584) for this purpose. Hacking may be seen as a form of citizens' data agency (Pybus et al., 2015), which may be employed for resistance as well as the explicit forwarding of political causes such as improvement of governance.

The alternative data and analytics practices developed and made known by the Quantified Self (QS) community exhibit a different take on civil society's possibilities of engagement with datafication. The community engages in 'soft resistance' (Nafus & Sherman, 2014) towards dominant data practices; it welcomes commercial actors, but it simultaneously questions who gets to aggregate data, how, and for what purposes. The community remains, therefore, ambiguous in terms of its

valuations, so that the values of sharing that are essential to the movement can thrive, but this can happen alongside the simultaneous commercialisation of self-tracking (Barta & Neff, 2016).

## 4.2 Data activism as a heuristic tool

As was pointed out above, data activism can be viewed as forming a response to the complications of straightforward notions of citizens simply deriving agency and empowerment from data technologies. Milan and Gutiérrez (2015) and Milan and van der Velden (2016) consider data activism as civil society's response to the uneven distribution of data and capabilities, indicating diverse social practices that manifest a critical attitude to datafication. Understood in this broad manner, data activism can include heterogeneous forms of activism and activist tactics. Data can be the issue of political struggle in their own right, or rather tools for struggle over other issues (Beraldo & Milan, 2019). Nevertheless, data activism starts from the recognition that as data and information are increasingly important in contemporary society, access to data is power. In data activism, citizen empowerment is seen to emerge from citizens' exercise of control over technologies (Milan & Gutiérrez, 2015; also see Rodriguez, 2001). In this way, data activism represents a potential challenge to existing data power relations. For data activism, "technology is simultaneously the means to provoke change in society and a site of struggle in its own right" (Milan & Gutiérrez, 2015: 127) due to the politics and power relations embedded in technology. Milan and van der Velden (2016) expand on these notions by considering data activism as civic engagement and political action that engage with new forms that knowledge production takes. In relation to dominant modes of datafication, data activism aims to create novel and alternative data arrangements and new responses to Zuboff's (2015) questions about who can learn based on data, and who decides about this. Data activism develops innovative ways of relating to datafication, the production of knowledge, and their consequences.

Given this broad characterisation of data activism, rather than a "definition" of a specific form of social action, Milan and colleagues consider data activism as a heuristic tool to explore how people politically engage with the production and use of data (Milan & Gutiérrez, 2015; Milan & van der Velden, 2016). Being a heuristic tool means that data activism is a concept that can enable understanding of a phenomenon; it is an artificial construct that acts as a lens through which to examine how activism evolves with respect to datafication. The concept is not even expected to be stable, but instead susceptible to revisions through empirical research (Milan & van der Velden, 2016).

### 4.3 Data activism and social justice

Kennedy (2018) explicitly ties data activism in with a notion of justice; data activism “seeks to challenge existing data power relations and to mobilise data in order to enhance social justice” (Kennedy, 2018: 18). In this view, data activists aim at the development of more just practices of data collection and use, as well as policies that govern such practices. These more just data arrangements, of course, need to involve normative notions of injustices to be addressed. In scholarship on data justice (Dencik et al., 2016; Taylor, 2017), social harms resulting from contemporary data arrangements are seen to exacerbate existing social injustices, as well as to produce new ones. Jonathan Cinnamon (2017) usefully analyses injustices produced by surveillance capitalism by framing them in terms of Nancy Fraser’s (2008) notion of parity of participation. This analysis can be employed to highlight different ways in which data activism can address injustices.

Fraser’s (2008) suggestion is to submit all justice claims to the normative principle that everyone should be permitted to participate as peers in social life. According to Fraser, injustices that hinder parity of participation in social life occur in three dimensions: they can concern the economic maldistribution of resources that prevents some from participating, socio-cultural misrecognition that affords some a lower status in the society, and political misrepresentation that denies political voice and possibilities to redress injustices. In Cinnamon’s (2017) analysis, maldistribution of data results from data extraction practices which separate people from their data. As a result of data extraction, data are accumulated by corporations that can then access and use data on an aggregate level, something that data subjects themselves cannot do. These mechanisms were already extensively discussed in Chapter 3. According to Cinnamon, data maldistribution can be seen as the initial injustice. After people are separated from their data, further injustices can be produced when those data are used by others in ways that have consequences on people’s lives. Socio-cultural misrecognition is one such further injustice (Cinnamon, 2017). The separation of people from their data makes it possible to employ data for profiling, social sorting and other forms of categorisation based on data analysis. The injustice of misrecognition occurs when people are categorised in ways that produce status inequalities that are consequential in their lives. The third dimension of injustice is intimately related to the second one; it concerns misrepresentation, or how people or groups are rendered voiceless to contest injustices and engage as peers in democratic society. According to Cinnamon, the mechanisms that propagate this injustice are related to the inability to meaningfully access data that people are initially separated from, as well as to contest the unjust categorisations done based on the data.

This analysis of surveillance capitalism’s injustices points out ways in which data activism can challenge injustices of data arrangements: by means of removing

obstacles that stand in the way of parity of participation in social life. Some forms of data activism can focus on data distribution, that is, address the injustice of data maldistribution. Others can focus on the consequences of maldistribution down the line, that is, socio-cultural misrecognition based on data, or political misrepresentation leading to the inability to challenge observed injustices. This, again, points to the various forms that data activism can take, as will be further discussed below.

## 4.4 Varieties of data activism

Paraphrasing Winner's (1986: 105) above-quoted disdain at how our thinking is driven by trust in computers and communication technologies, data activism can be thought to be premised on contesting the almost religious conviction that datafication will automatically lead to a better world for human living. However, even drastic differences between forms of data activism are visible in how this contestation unfolds. One issue is whether datafication is seen as something that needs to be avoided, or alternatively as a potentially positive force that needs reorientation. Starting from one of the first theoretical outlines of data activism (Milan & Gutiérrez, 2015), this distinction between attitudes toward datafication has been noted in two "ideal types" of data activism: reactive and proactive data activism (also see Beraldo & Milan, 2019; Milan & van der Velden, 2016).

### Reactive data activism

Reactive tactics posit the extraction, analysis and monetisation of data as a threat to individual rights, and consider datafication as something dubious, a thing to be avoided. Considered as a response to the injustices arising from dominant data practices, reactive data activism is focused on data distribution. It attempts to prevent the production of data, and therefore also the injustice of data maldistribution, in the first place. While the aim is to avoid potential harms, this avoidance also means bypassing any positive outcomes associated with data use. Reactive data activism involves self-protection against dataveillance through technical means: anonymity, obfuscation and encryption (Milan & van der Velden, 2016). On the practical level, this may include employing and promoting secure online practices, such as the TOR browser, using tracking protection in web browsers or VPN tunnelling in the internet connection, email encryption protocols such as GPG or PGP, or end-to-end encrypted messaging software such as Signal. Examples that can be viewed as reactive data activism also include employing alternative platforms such as DuckDuckGo for search and Mastodon for social networking, as well as general promotion of open source alternatives to popular web services. Abstaining from



Facebook use or removing mobile apps (Perrin, 2018) in the wake of Cambridge Analytics revelations in 2018 also fall in this category. Besides self-protection, organising events to help people with safer online communication (Kannengießer, 2019) or urging people to remove their Facebook accounts (Biggs, 2018) can also be examples of reactive data activism.

Reactive responses are largely efforts to intervene in social problems on the level of individual behavioural change and, as alternatives to dominant data arrangements, are dependent on acts of individual resistance (Dencik, 2018). Reactive data activism focused on only one service or platform also risks ignoring the broader political economy of dataveillance. Its more technical modes are reliant on individuals' technical skills and know-how, and as such have not been successful at moving beyond expert communities (Dencik et al., 2016). Abstention from popular search, communications and social networking platforms are essentially acts of individual media refusal (Portwood-Stacer, 2013) and as such are a limited tactic of political engagement. Platforms for messaging or networking, as well as online search, recommender systems and other services reliant on large amounts of user data, are also subject to network effects which makes less-popular alternatives also less valuable to use. More generally, even if reactive data activism is empowering for individuals who engage in what they consider as privacy-preserving activities, "individual actions infrequently aggregate to facilitate changes in industrial infrastructure that result in collective empowerment or systemic change" (Draper & Turow, 2019: 10) due to their failure to undermine powerful systems.

### Proactive data activism

Where reactive data activism starts from the conviction that datafication is ultimately harmful – at least given the political economy of data in which it is firmly embedded – and therefore the best thing to do is to avoid it, proactive data activism, the other archetype mentioned above, starts from a more positive outlook for datafication. Rather than resisting datafication, or even only attempting to mitigate its harms, proactive initiatives consider datafication as a potentially positive force. Proactive data activism views datafication as "an unprecedented, powerful opportunity to provoke social change" (Milan & van der Velden, 2016: 67) and "sees people's active engagement with technologies a pathway to empowerment, equal participation and action" (Milan & Gutiérrez, 2018: 58). The focus is on taking advantage of data, data technologies or data infrastructures, and on appropriating them for advocacy goals to serve data activists' ideas of desirable change.

A further way to make sense of differences between proactive forms of data activism is to analytically distinguish between data as a repertoire of action, and data as an issue at stake (Beraldo & Milan, 2019). The first instance, data as a repertoire,

is evidenced by proactive initiatives that make use of data by processing, analysing, visualising and crunching them, for the explicit purpose of furthering some well-defined social ends or specific advocacy goals. Milan and Gutiérrez discuss data activism as a field of action that emerges at the intersection of communicative processes in the form of different kinds of media activism, and information-related professions such as data analytics and journalistic investigation. They stress the significance of technology to manipulate data, and consider data activists as “interpreters of data, acting as facilitators in the contemporary data-rich public sphere” (Milan & Gutiérrez, 2015: 126). This viewpoint to data activism stresses the aim of making data serve concrete social aims; data activism is about making data serve the ends of advocacy. Depending on the case, gaining access to the data may involve different kinds of activism, such as whistle-blowers, employing open data, crowdsourcing data, appropriating existing data or creating data (Gutiérrez, 2018a). Concrete examples include attempts to impede environmental threats through the promotion of data transparency, data collection, data sharing and visualisation (Milan & Gutiérrez, 2018); and using crowdsourced crisis data for humanitarian assistance, effectively creating new spaces for debate and action (Gutiérrez, 2018a; 2018b). The use of citizen and civil society data projects as advocacy instruments to affect data collection by official institutions on issues such as killings by police, land registries, water supply and literacy (Gray et al., 2016) also fall in the category. Further examples in connection to citizen engagement include civic hackers who request, digest, contribute to, model, and contest data, and make use of open data in an attempt to guard the public against injustices (Schrock, 2016). Stemming from how these proactive initiatives and projects employing data for social change ends have defined their goals, they have specific targets in correcting the injustices of misrecognition and misrepresentation.

In contrast, some proactive data activism initiatives are rather oriented at considering data as the issue at stake (Beraldo & Milan, 2019). Open data activism provides one example. While some open data activists may also have a more specific social change goal in mind, in which open data itself rather becomes a repertoire of action, open data’s general premise is that data produced by public authorities should be technically and legally free to use, distribute and reuse (Baack, 2015; Kitchin, 2014). Open data falls under the heuristic of data activism. It has emerged as a reaction to how datafication has resulted in an uneven distribution of power and knowledge that favours companies and governments and hinders public agency (Baack, 2015). In terms of challenging the unjust distribution of data, this premise is grounded on the belief that positive societal and economic transformations, unknowable at the moment, will take place once data maldistribution is attended to. Open data activists explicitly identify data maldistribution as a problem by regarding the availability of raw data as a prerequisite for generating new knowledge, so that

the monopolisation of data leads to the monopolisation of knowledge production (Baack, 2015). Consequently, open data activists attempt to balance the distribution of power and knowledge in society by turning data from a proprietary resource held by governments into a public resource benefiting the society at large. They are committed to the general aim of democratising data and knowledge production, rather than a set social change objective that can be supported by a specific kind of data manipulation. Aided by open data policies, others are then expected to access and make use of open data to forward specific social change objectives. In order for this to be possible, new kinds of technical infrastructure are required that make opening data possible (Kitchin 2014, 57). In addition, open data activists see the necessity of developing intermediaries, including new technologies, that make knowledge production possible (Baack, 2015). If infrastructure is considered as technical forms that make the flows of other things possible (see Larkin, 2013), open data activism could be considered an “infrastructural” form of data activism, as it is about the enablement of data arrangements that make movements of data possible, and those movements can be harnessed by others to further their own ends.

Considering these rather different forms of civic engagement that may be regarded as proactive data activism, looking at the phenomenon from different viewpoints and focusing on different empirical cases will highlight a variety of aspects. Miren Gutiérrez (2018a: 49–106) discusses how the phenomenon can be theoretically viewed through the lenses of, for example, communicative action, data journalism, citizen media, social movement studies, citizen monitoring in the public sphere or technology activism. Gutiérrez then empirically classifies cases of proactive data activism as, for example, skills transferrers who in essence make data activism of others possible, producers of tools and platforms for data activism, catalysts who provide financial and other resources for data activism projects, producers of data journalism content and visualisations, and campaigners and advocates. An empirical data activism case can exhibit a combination of these attributes, and any individual data activists may engage in more than one of the modes of action.

## 4.5 Data activism and the private sector

The relationship to the private sector can be another issue differentiating between data activism initiatives. In some cases of data activism, industry involvement has been considered as inherently dubious and suspect. Reactive data activism focus on avoidance of data extraction exhibits an adversarial attitude towards the private sector and, in the post-Snowden era, towards the state-industry complex (Dencik, 2018; Milan & van der Velden, 2016). For other initiatives, the relationship with the private sector may be more nuanced. Open data, for example, can support liberal

democratic values by providing mechanisms for more just governance, but also libertarian agendas by providing justification for privatisation and deregulation (Schrock, 2016). Analysing open government data initiatives, Bates (2013) highlights how political processes can restrict the counter-hegemonic potential of open data and instead shape it to support the marketisation of public services. Critics, including both practitioners as well as researchers, have more broadly objected to the involvement of corporations in open data, citing concerns over dubious political alignments and the potential for co-optation (Schrock, 2016). In a similar vein, Johnson (2014) points out that open data exists only in relation to the political economy and socio-technical arrangements, and instead of promoting more just data arrangements it may lead to further inequities. This is, in part, due to the asymmetries between capabilities to do things with data. In practice, data may be open to commercial enterprises, but not citizens (Johnson, 2014). Research has, however, pointed out that data activists are cognisant of some of these issues and the factors underpinning them (Baack, 2015; Schrock, 2016). Nevertheless, these critical views on open data highlight a broader concern for data activism; when activism agendas and commercial realities are in tension with one another, activism focusing on data distribution may serve agendas and imaginaries of other beyond the activists themselves.

Milan and van der Velden (2016) point out also a conceptually different form of private sector involvement. If data activism promotes alternative technologies and policies associated with them, it may involve some form of collaboration with private sector actors. This notion is based on the concept of technology- and product-oriented movements, or TPMs, discussed by David Hess (2005): mobilizations of civil society organizations that include alliances with private-sector firms. According to Hess, such collaboration can serve pragmatic ends; when data activism's goals require the development and production of alternative technologies, it may resort to private sector firms to provide them. Firms can be motivated to collaborate with data activists in order to seek opportunities such as securing or expanding markets for their outputs. In its attempts to mobilise new data arrangements, data activism may therefore concern firms as participants and beneficiaries.

## 4.6 MyData as data activism

The discussion above points at the conclusion that data activism is an umbrella concept, covering a wide array of critically-oriented approaches to datafication (see Gutiérrez & Milan, 2019). Here, datafication is understood as a societal phenomenon, so that a critical approach to datafication does not necessarily mean being critical of the technical processes of transforming aspects of people's lives into quantified data *per se*. Instead, data activism can approach datafication from the

point of view of not only technical but also political, philosophical, ethical or economic considerations. As discussed above, the field of data activism can imply different relationships to data economy's existing arrangements such as the commercial data use, institutions regulating such use, as well as their governance. Different ways to categorise data activism were identified above, including reactive and proactive data activism; data as action repertoire for activism and data as stakes of activism; the different social injustices that data activism aims to resolve; as well as data activism's relationship to the private sector. These categories are analytical rather than empirical (Beraldo & Milan, 2019); for example, some open data activists may consider data as both stakes and a repertoire, lobbying general open data practices for the purpose of aiming at distinct goals by means of consequently opened data. Rather than a strict definition to help with delineating whether or not something is data activism, I follow Milan & van der Velden (2016) and consider data activism as a heuristic to make sense of forms of engagement and political action spurred by datafication.

As discussed in Chapter 1, the Finnish-originated data governance initiative, MyData, forms the empirical context of this thesis. While the initiative is not necessarily critical towards datafication as such, it is critical towards the political-economic aspects of data governance and control of data; that is, its goals concern who can learn from data and who decides about this. Keeping well in mind that the different categories of data activism are analytical rather than empirical, if we make use of the heuristic to understand MyData, what does the initiative look like? How could we position MyData in the space of data activism spanned by the analytical categories?

On the spectrum of reactive and proactive data activism, MyData falls in the proactive end, in the sense that datafication is seen as a potentially positive force, an opportunity to provoke social change. While the belief is that desirable outcomes for people, businesses, and the society at large will emerge once people are in control of how data are used, the primary aim is on the level of data governance rather than on any specific desirable outcomes. The initiative is oriented towards data, and rather than being an action repertoire that enables promoting other activist goals, data themselves are what is at stake. MyData can be considered as an infrastructural form of data activism in a similar sense as open data. The goal is to set up new data arrangements that make possible the movement of data under new principles, and these arrangements, then, may be employed by others to further their own goals.

Viewing data activism as promotion of more just data arrangements, MyData can be distinguished via the injustices it aims to resolve. This issue is discussed in Article III. The primary injustice is data maldistribution, the unjust distribution of data that denies people the resources to participate in the digital environment on an equal footing. With the goal of placing people in control of the uses of their data,

MyData is focused on releasing data from a proprietary regime to new uses, and therefore redistributing data, and benefits derived from them, from data-gathering organisations to people. At the same time, the aim is to resolve another form of maldistribution, that is, the uneven distribution of personal data between those firms and organizations that are positioned as primary data collectors, and their less data-endowed counterparts.

In order to achieve its aim to increase people's capabilities to act in relation to their data, MyData collaborates with private sector actors producing data-related technologies and services, some of which are explored in Article II. This collaboration highlights MyData's likeness to TPMs (Hess, 2005), where collaboration serves both the ends of a social movement advocating alternative technologies and policies associated with them, as well as the ends of private sector firms producing these technologies. With an eye on serving firms' economic interests in personal data, the initiative explicitly aligns its interests with commercial data use; as proclaimed in one white paper, "[MyData] combines digital human rights and industry need to have access to data" (Poikola et al., 2015: 4). Having individuals control their personal data is expected to make possible also new kinds of commercial uses of those data, a dual aim explored in Articles III and IV.

Given how it has from the early stages been involved in Finnish policy agendas, and given the role of private sector in MyData-related technology development, the initiative could obviously be approached with other lenses in addition to, or in place of, data activism, such as with more focus on the field defined by either technology-related policy (i.e., the state) or technological entrepreneurship and business (i.e., the market). Here, a comparison to a better-known example of data activism, open data, can provide insight. Like MyData, open data involves different stakeholders with different goals. On one level, open data refers to a specific model of data governance, and data can be said to be "open" if it is subject to this governance model (Kitchin, 2014). The issue at stake can be the promotion of open data as a general principle of organising public sector data arrangements (e.g., Baack, 2015). Some data activists leverage open data principles in order to access certain datasets to forward specific activist ends (Schrock, 2016). Open data is also a policy implemented by different government actors within their own data arrangements (e.g., Kitchin, 2014: 50). Simultaneously, open data allows a role for private sector actors in making use of the data becoming available as a result of open data policies (e.g., Bates, 2013; Johnson, 2014; Kitchin, 2014). As a result, open data is something that can be useful for the forwarding of varied ideological ends (e.g., Bates, 2013; Schrock, 2016). In a similar way, MyData could be approached in many ways; for instance, as a budding technology or competition policy concept, a definition for a data governance model or for a certain subset of personal data, or as a means of defining a potentially emerging field of data business. Noting these other possibilities, which might be

pursued elsewhere, I employ the heuristic of data activism to investigate the MyData initiative and the technology development related to its goals. This means focusing on MyData's aims to affect citizens' capabilities to act in relation to their personal data, and to have a say in how their personal data are employed for production of economic value.

# 5 Research Approach

This thesis is based on research originally published in Articles I–IV. Each of the articles had a distinct research approach and research questions. Taken together the articles contribute towards two research aims of the thesis, which are reflected in three central research questions.

The first research aim is to investigate what data activism aiming to have people control their personal data is about, and to understand its political and ideological underpinnings. The first two research questions reflect this aim. These questions are framed in terms of the literature on collective imagination. The point of departure is that multiple imaginaries about the data economy exist, potentially in tension with one another, or possibly in a more dialectical relationship. In Chapter 3, I outlined the dominant imaginary about data economy’s data arrangements based on data studies literature, highlighting particularly the position imagined for citizens and consumers. In Chapter 4, I discussed literature on data activism, considering it as a form of civic engagement producing alternative imaginaries for the data economy. In light of the dominant data economy imaginary and alternative imaginaries produced in data activism, the data economy can be viewed as a field of contestation over the collective imagination. The first research question and its sub-questions address the interplay of dominant and alternative imaginaries:

1. How do alternative data economy imaginaries and the dominant imaginary compare with each other?
  - a. What new positions are imagined for citizens and consumers in the alternatives?
  - b. How is this repositioning imagined to intervene in data economy’s economic logic?

In light of the discussion in Chapters 2 and 4, the politics of imagination concerns also multiple alternative imaginaries that can be developed in data activism. There are two levels of contestation that I am interested in: data economy as a field of contestation over the collective imagination, and data activism as a field of contestation over the alternative imaginary. The latter level is reflected in the second research question and its sub-questions:



2. How do different alternative imaginaries developed in data activism compare with each other?
  - a. What tensions emerge between alternative imaginaries developed in data activism?
  - b. What affects the success of alternative imaginaries, that is, their expansion outside data activism?

The second research aim of this thesis is to develop, through the understanding of the empirical phenomenon, a position on the imagined alternative data arrangements. From a standpoint informed by data studies scholarship reviewed in Chapter 3, datafication and the dominant data arrangements in the data economy have far-reaching consequences that deserve to be addressed. In light of this literature, it seems justified that the researcher, in addition to observing attempts to shape different pathways for the data economy in a detached mode, becomes involved in those attempts in a more engaged mode. As Schrock (2016: 583) points out, data activism represents a moment where meaningful change may occur, and we should be attentive of these moments. From the standpoint informed by the data studies scholarship, however, it would be unjustified to engage with data activism in a manner that simply accepts the agenda, the problem settings and the normative notions of data activists, and therefore risks under-problematising them. An observer of data activism practice does not require particularly honed critical faculties to identify problematic aspects in the imagined rearrangements of the digital environment. The second aim of this thesis is to approach data activism in a way that does not only critically pick apart the developed imaginaries and does not only engage in data activism practice either. Rather, the aim is to open up a space for the exploration of its notions of desirable data futures. This research aim is normative, but rather than committing to a predetermined agenda, the aim is to produce normativity in relation to data activism's aims and approaches. This is reflected in the final research question and two more specific sub-questions:

3. How can we identify and promote societally desirable data economy imaginaries?
  - a. Which alternative data arrangements may be considered desirable?
  - b. How can desirable data economy imaginaries be promoted in data activism?

## 5.1 Research approach of the original publications

In this section, I provide an overview of the research approach taken in Articles I-IV, including the research questions, research setup, empirical data, and methods. Summary of the research approach, along with key findings, is provided in Table 2. A more detailed discussion of the articles' findings will follow in Chapter 6.

Articles I and II are outcomes of mapping of a phenomenon emerging under various labels during the research process. The articles concern new technologies that rely on, in a fundamental sense, a similar understanding of problems, their possible solutions, and the desired economic and social order that was to result from this technological intervention. In terms of the broader research aims of this thesis, they provide insight into the field under study, as well as critical insight into the imagined alternative data arrangements. In Articles III and IV, the focus was shifted from investigating the developed or imagined technologies, to empirically investigating MyData, which was gaining momentum as a data activism movement working towards social change. In connection with the research aims of this thesis, the articles document and analyse the data economy imaginaries underpinning MyData. Article IV also details one attempt to engage with data activism in a way that is both critical and constructive.

### Article I: Consent intermediaries

“Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach” was co-authored with Yki Kortensniemi and published in *Big Data & Society*. The research was motivated by the identification of a recurring concept in ongoing initiatives aiming to have people control their data. Our research project had done a mapping of emerging services going by names such as “personal data management platforms”, “personal information management systems”, “MyData operators” or “personal data spaces”. While their practical implementations and stages of maturity varied, we found that these services could be viewed as middlemen of consent decisions. They offered to bring the decisions under one control point through which individuals grant, view and withdraw permissions to collect and use data. In Article I, we called these services by the moniker consent intermediaries, or CIs. As discussed in Chapter 3, privacy self-management relies on individuals being informed and making decisions based on subjective analysis of this information, but such analysis is anything but straightforward in the contemporary context of the data economy. We asked whether, how, and to what extent such a CI could aid with the challenges posed to privacy self-management. At the outset, an approach based on introducing a new intermediary technology appears to rely on a technological solution to a problem

**Table 2.** Summary of the research approach and findings of Articles I–IV.

	<b>Article I</b> Consent intermediaries	<b>Article II</b> Personal data spaces	<b>Article III</b> Data agency at stake	<b>Article IV</b> The social imaginaries of data activism
<b>Literature</b>	Privacy research	Data studies	Data activism research, data justice	Data activism research, social imaginaries, data studies
<b>Data</b>	Published research literature	Interviews, questionnaire responses	Keynote talks, audience interaction data	Field notes, interviews, reports, memos, social media posts, etc
<b>Methods</b>	Review of literature, conceptual analysis	Qualitative analysis based on open coding	Qualitative analysis of collective action frames	Participant-observation
<b>Approach</b>	Engaged in terms of problem setting	Critical in terms of problem setting	Critical in terms of problem setting	Critically engaged to produce normativity
<b>Key findings</b>	<ul style="list-style-type: none"> <li>• Informed controlling of personal data is difficult under dominant data arrangements</li> <li>• An intermediary service can provide some, but limited, control aides</li> <li>• Difficulties partially due to the individualised model of data control</li> <li>• The intermediary itself gains power to affect data control decisions</li> </ul>	<ul style="list-style-type: none"> <li>• PDSs aim to provide users with capacities to act in relation to personal data</li> <li>• These are imagined to intervene in dominant data arrangements</li> <li>• Intervention is based in the reorientation of data markets</li> <li>• Users are imagined as capable participants in the data economy</li> </ul>	<ul style="list-style-type: none"> <li>• Lacking data agency framed as the primary obstacle to participation</li> <li>• Data agency to resolve injustices for both people and firms</li> <li>• Participation framed either as market participation or citizenship more broadly</li> <li>• Data activism's collaboration with firms favours the market participation frame</li> </ul>	<ul style="list-style-type: none"> <li>• Identifies alternative bases for data economy imaginaries</li> <li>• Imaginaries based on different ideas about the outcomes of data control</li> <li>• Combines the imaginaries for collective forms of data governance</li> <li>• Suggests a constructive role for critical scholarship in data activism</li> </ul>

whose roots lie in the cost-benefit based model itself. However, as we write in Article I, “for the moment, we will live with the model of informed consent, as in many jurisdictions it is codified in legislation” (p. 2). Further, it seems reasonable to assume that some technologies can better help users cope with the decisions they face. In addition, apart from CIs, also other practical suggestions to improve consent processes have been presented, such as expiry dates for consents (e.g., Custers, 2016), calling for further discussion on implementing consent in datafied times. This motivated us to explore this technological solution encountered in the empirical field to analyse its merits, limited as they might be, to a problem that is to some extent technical, but at its core more fundamental. We therefore explored the potential of CIs in a positive nature.

Article I can be described as conceptual research where the “focus is on integration and proposing new relationships among constructs” so that “the onus is on developing logical and complete arguments for associations rather than testing them empirically” (Gilson & Goldberg, 2015: 127). It did not aim to develop new theory, nor did it engage in the analysis of empirical data, apart from focusing on a concept identified in the empirical field of study. The research was based on observation and analysis of already available information, i.e. research literature, with the aim of understanding the merits and deficiencies of a proposed alternative approach to implementing privacy self-management. The method of analysis was the categorisation of obstacles identified in existing literature, and then reflecting the proposed solution against the categories.

## Article II: Personal data spaces

“Personal data spaces: An intervention in surveillance capitalism?” was authored by myself and published in *Surveillance & Society*. Personal data spaces, or PDSs, are intermediary services that promise to empower people to take control of processing of personal data by providing their users with a data storage service coupled with data management interfaces. Article II focused on three exemplars of such services: the “personal cloud server” Cozy Cloud, the “digital life management system” Meeco, and the “personal data store” OpenPDS. The first two are commercial products of two relatively small start-up companies; the third is a spinoff of a research project at the MIT media lab. The start-up firms remain up-and-running in 2019, while the research spinoff has online presence but seems to be discontinued or in hibernation. The developers of the three analysed exemplars have varying practical solutions for data storage and sharing. Despite this, between them they exhibit a common belief that people should be able to exercise more control over their data, and that this would lead to valuable outcomes for the people involved. Common beliefs also include the idea that there needs to be an intermediary service

through which control of data by the user becomes possible. PDSs seem to propose a potential intervention in dominant data practices by promoting new capacities for their users to act towards data. In the article, the three exemplary PDSs are approached as representations of an alternative imaginaries for the data economy. The point of interest were the interventions their developers were attempting to achieve to the economic logic currently dominating the data economy. I delineate these social imaginaries by first asking what agency towards data PDSs offer people. I then compare this imagined agency to the dominant data arrangements as identified in data studies literature. The aim is to examine how these services propose to transform the economic role of people in value creation from personal data.

Empirically, Article II was based on explorative interviews with three PDS developers, the developers' open-ended responses to a policy questionnaire collected by the European Commission, informal discussions with numerous PDS developers, personal use experience of two PDS services that were publicly available for use, and information available in public sources including the websites of the PDS services, instruction documents, media appearances of PDS developers, and a research publication on one of the services available at the time (de Montjoye et al., 2014). Data for the article were collected in 2015 and 2016. The interviews were specifically set up for research purpose, so that the interviewees agreed to participate in this research, and to the recording of the interviews. An external service provider was used to transcribe the recordings. The questionnaire material was gathered by the European Commission for the primary purpose of providing background information for a roundtable discussion. A report based on this data has been published by the Commission (European Commission, 2016). The Commission gathered the material with the pretext that results will be shared with participants of the roundtable, i.e. potential commercial competitors. Research use of the questionnaire responses was secondary; the Commission personnel shared the material for research purposes with the members of a research project I was involved in, and members of the project also took part in the roundtable. The analysis was initially guided by a broadly defined interest: to identify how the services envisioned features and connections to the broader data economy. The publicly available material, use experience of the services, and informal discussions made up background material that both guided the analysis as well as informed the case descriptions presented in the article. During the course of analysing the interview transcripts and questionnaire responses, I further focused to the ways that the services were imagined to afford their users agency over data. This part of the analysis was based on open coding using *Atlas.ti*. The materials were coded with a focus on what end-users were doing, or imagined to be doing, with the PDS. The coding was iterated, so that in the end the four aspects of imagined agency for PDS users, as presented in the article, were reached.

### Article III: Data agency at stake

“Data agency at stake: MyData activism and alternative frames of equal participation” was co-authored with Jesse Haapoja and published in *New Media & Society*. In the context of this thesis, the article switches focus from the technologies under development to data activism. The article takes as its starting point the notion that data activism attempts to shape and mobilise more just data arrangements. We build on the notion that justice requires arrangements permitting all to participate as peers in social life. Data activism, in this view, is about identifying and removing obstacles to equal participation in the digital environment. MyData activism, similarly to some other forms of data activism, involves private sector actors to achieve, in practice, the imagined more just data arrangements. Firms involved in data activism, in turn, seek policy and market support for their products and services. Here, data activism becomes to concern firms as participants and beneficiaries. In article III we analysed collective action frames constructed in the first MyData conference, which became a formative event for the MyData movement. In the conference, actors including data activists, firms, and policymakers attempted to shape MyData to suit their activist, commercial, or policy ends. In this context, a professional conference is a venue for contestation between different future visions, and an environment of selection between alternatives. The collective action frames inform us how injustices and their remedies are framed in the context of data activism that involves commercial actors as participants. In the article we asked what injustices hamper equal participation, what are their remedies, and whose interests deserve consideration.

The article is based on empirical data collected at the three-day conference organised in August 2016. Having participated in a thematically related research project, we were involved in a minor task in conference organisation, arranging a workshop aimed at academic researchers. This, and our general proximity with the conference organisers, allowed the collection of two datasets during keynotes. The first dataset consists of transcriptions of video recordings of 12 keynote talks and their follow-up discussions, totalling some seven hours of professionally produced video material. The videos were transcribed for analysis by an external transcription service provider. The second dataset consists of about 750 anonymous, mostly tweet-size messages sent by the conference audience members by means of a backchannel software (Nelimarkka et al., 2016). The software allowed audience members to use their mobile devices to send questions, comments and specifically requested ‘lessons learned’ during keynotes. Video recordings of keynote presentations are also publicly available. Compared to our dataset, the published recordings exclude follow-up discussions and responses to questions asked by the audience. All messages sent via the backchannel software were anonymously public during the

event, and from the software logs we likewise received only anonymous data. No confidential data was handled during data collection and analysis.

As an analytical framework for keynote presentations, we employed the identification of collective action frames (Benford & Snow, 2000). Frames in general offer a schema for highlighting certain aspects of a situation, functioning as modes for articulating strategy to be undertaken. Following Snow and Benford (1988), collective action frames diagnose the issue in need of change and who is to blame, prognose solutions and how to achieve them, and motivate collective action. Using *Atlas.ti*, we initially identified sections that represented collective action frames and concerned participation in the data economy. Both authors first separately identified and classified these sections with an open coding scheme, and we then collaboratively and iteratively reclassified them until reaching the six frames presented in the article. We included in the analysis only frames that were either widespread among the keynotes, or that were contested. The audience interactions allow investigating the success and reception of framing efforts (Snow & Benford, 1988). For this purpose, we identified agreement and tension arising in response to frames identified from the keynotes. This means that the analysis was first performed on the keynote talks, and when we had identified the frames employed in the talks, we analysed audience comments in relation to these frames.

#### Article IV: The social imaginaries of data activism

“The social imaginaries of data activism” was co-authored with Minna Ruckenstein and published in *Big Data & Society*. The article employs the framework of social imaginaries to investigate MyData. In contrast to Article III that was focused on a single formative moment, Article IV is based on participant-observation with data activists over the longer term. Studying an initiative such as MyData means dealing with a work-in-progress and uncertain futures in the making; this meant that in our research we engaged in an ongoing observation and dialogue when interacting with MyData activists. Via this engagement, we realised that data activism is about alternative futures in the making in two senses. It is not only that activists working in the initiative envision data futures as alternatives to the future they view as problematic, but that we face alternative data future visions also within the initiative. The starting point was that alternative social imaginaries can coexist simultaneously in the context of one data activism initiative. From here we had two aims. The first aim was to unpack different future imaginaries that are present in data activism, aiming to clarify the political and social alternatives that different social imaginaries ascribe to the notions underlying data activism. In practical terms, we do this by outlining two alternative social imaginaries at play; one that was represented by technology developers and one that we ourselves, as social scientist, represented.

The second aim was to produce insights that can be employed for re-articulating the aims of data activism, attempting to rework the different imaginaries into a shared dialogue. Towards this end, we discuss how to merge the two imaginaries in the work of reimagining data-related governance structures and knowledge practices.

Empirically, Article IV draws from longer-term participant-observation and working together with developers and data activists between 2014 and 2018. The main context for this participation were three research projects in the fields of personal data, health and knowledge work, and the participation consisted of countless everyday interactions, discussions at project meetings, and formal and informal interviews. Alongside activities in the research projects, we took on participant-observer roles in a 450-member (in Autumn 2018) Facebook discussion group called “MyData working group”. The group consists of civil servants, activists, technology developers and start-up entrepreneurs. During the course of four years, I also participated in the Finnish MyData industry alliance, where a national MyData model was being developed through pilot projects. Further, we did fieldwork in our roles as organisers, presenters and observers at three international MyData conferences in Helsinki in 2016, 2017 and 2018. The empirical materials, therefore, consist of material produced during the participation, including meeting presentations, field notes, emails, social media posts, reports, white papers and plans, published documents referenced in the article. The empirical work done for Articles II and III was part of the same longer-term research process, and their empirical data also informed Article IV.

## 5.2 Research approach of the thesis

The order in which the articles appear in this thesis corresponds to the chronological order of conceptualising and carrying out the research work underlying them. The order also corresponds to the breadth of the problem setting. Article I addresses a specific problem that is motivated by emerging intermediary technologies. The problem setting is the most specific one in the sense that it is laid out in terms of privacy literature, and connected the relatively narrow understanding of the issue at hand in terms of privacy management. Article II continues with a partially overlapping problem setting by continuing to examine intermediary technologies. However, instead of understanding the issue at hand in terms of privacy, it broadens the view to encompass data economy’s economic logic. Article III again broadens the view, continuing with the economic logic but expanding from intermediary technologies to considering how problems and their solutions are framed during the formation of a data activism movement. Finally, Article IV provides the broadest view by focusing on the different social imaginaries underpinning data activist work, and it can be viewed as an expansion of the notion of alternative frames in Article



III. Considered via the research questions of this thesis, the problem settings of the articles are partially overlapping, with each subsequent article covering parts of the problem setting of the previous articles.

The insight gained during this process of expanding the problem setting has not only increased scholarly knowledge about the empirical field, but has also informed different modes of relating to the empirical field and the normative commitments encountered in the field – in this sense, the articles trace a path of alternating between different modes of this researcher’s engagement in the field. Here I do not intend to claim that the specific modes of engagement taken in the course of the research would have been planned when the work on this thesis was started. Rather, the point is that during the research, I started to understand that the “messy” (Law, 2014), dynamic and at times slippery nature of the object of research necessitated assuming different positions in relation to it. At times, the object of research seemed to change as soon as an approach was taken. It seemed to me as if I was looking at one of the inkblot cards of the Rorschach test; if I would show it to someone else, or look at it at a different time, it would transform into something else altogether. Is the empirical phenomenon about privacy in the digital environment, as the approach taken in Article I suggests? Yes, but it is not about privacy only. Is it about market agency, as the findings of Article II suggest? Not only, and especially not always, and not to everyone, as Article III shows. Further, the vision about the future that brings different actors together in data activism allows for a considerable amount of interpretive flexibility, as discussed in Articles III and IV, also making it possible for the actors to come together in the first place.

The empirical phenomenon guided the research strategy of alternating between modes of engagement, and also guided the choice of those modes in the course of empirical work. I argue that this has made it possible to produce more insightful knowledge about the field of data activism. There is also another point to make about engagement. As John Law writes about objects of social scientific interest, “since social (and natural) science investigations interfere with the world, in one way or another they always make a difference, politically and otherwise. Things change as a result. The issue, then, is not to seek disengagement but rather with how to engage. It is about how to make good differences in circumstances where reality is both unknowable and generative” (Law, 2004: 7). The modes of engagement taken in the course of this research reflect my modest attempt at seeking research methods that can, as Law suggests, “imagine and participate in politics and other forms of the good in novel and creative ways” (Law, 2004: 9) and that can help, in addition to learning more about realities, also “participate in the *making* of those realities” (Law, 2004: 10, original emphasis).

The relationship between the researcher and the researched is a topic of long-standing discussion in social sciences. For example, one of Karl Marx’s *Theses on*

*Feuerbach* proclaimed that the point of scholarly work was not to interpret the world but to change it; Max Weber, in contrast, while maintaining that social sciences are necessarily value-laden, argued in the lecture *Science as a Vocation* that "the prophet and the demagogue" do not belong on the academic platform (Zuiderent-Jerak, 2015: 1; 5–6; also see Reiss & Sprenger, 2017). According to these two potentially oppositional positions, the scholar should either aim at actualising an explicit political position through scholarly work (in Marx's case, this political position would be related to correcting a particular power dynamic observed in the society); or should avoid furthering utilitarian and political ends in their work (in Weber's case, this was a matter of trying to abstain from furthering political ideas as far as possible).

The former of the two corresponds to a specific "critical" mode of scholarly work. In the critical mode referred to here, there is a practical and explicitly normative goal and commitment to identify and overcome dominant power dynamics, providing "bases for social inquiry aimed at decreasing domination and increasing freedom in all their forms" (Bohman, 2016). Theories that are in this sense critical have emerged, for example, in connection to social movements that identify and oppose forms of domination or oppression of human beings in the modern societies. In the field of scholarship related to the data economy, particularly surveillance studies and data studies can have this orientation. As I have drawn extensively from these scholarly fields, a certain philosophical orientation towards this mode of scholarship is embedded in the approach of this thesis.

The latter of the two is closely related to the view that objectivity is an ideal that scientific inquiry should strive for. At least some views of scientific objectivity seem at odds with the critical mode of scholarly work; however, it should be noted that the goal of objectivity can be conceived of in several ways (Reiss & Sprenger, 2017). Objectivity can be seen to imply faithfulness to facts "out there", or to require that personal biases are absent from scientific observation and reasoning. Alternatively, objectivity can be taken to mean that scientific claims and practices are objective only to the extent they are free of moral, political and social values. A critical orientation in the above sense should not, for example, preclude objectivity in the sense of faithfulness to facts, nor should it be a reason to give free reign to personal biases when interpreting results. It does, however, mean that this scholarship is indeed not free of moral, political or social values, as such values are explicitly built-in in the research itself. However, there are good grounds to ask whether any research, particularly given the object of research in the social sciences, can be fully free of moral, political or social values.

One means of navigating between critical and objective modes of scholarly work is engaged scholarly practice that is often understood as taking an "insider" perspective in the empirical field of study. Engagement is presented as a way to go

beyond the dualism between critique and objectivism, and to produce scholarship that is relevant to the field, but at the same time scientifically rigorous (Zuiderent-Jerak, 2015). During the work towards this thesis, there was a strong push, or perhaps rather a pull, towards a more engaged researcher position. In the multi-disciplinary MyData-related research projects that I participated in, both the project requirements as well as the data activism practitioners expected the social scientist to not be a detached observer, but to participate in and constructively engage with the empirical field. Such requests for constructive participation echo demand for design input from social scientists long experienced in the fields of human-computer interaction and systems design (e.g., Anderson, 1994; Hughes et al., 1994). This is not necessarily problematic. Like data studies scholarship, data activists recognise the far-reaching consequences of datafication, which enables new approaches for making sense of the world that in turn affect the production of knowledge, business practices and governance. This would indicate that it can be in alignment with scholarly commitments to engage in the data activism field and to examine its critical potential to rework and re-imagine processes related to datafication. At the same time, a critical commitment makes it necessary to turn the analytical apparatus also towards the imagined alternatives. Approached in a too straightforward manner, engagement in data activism practice can mean contributing scholarly knowledge to pre-set problem definitions as they are encountered in the empirical field. This kind of engagement produces a legitimacy problem for a researcher with a “scholarly attachment to complicating underlying assumptions” (Zuiderent-Jerak, 2015: 8). It can lead to a limited variety of intellectual positions that the researcher can assume, as well as to delimiting the scholarly imagination when it comes to relations with the empirical field (see Jensen & Lauritsen, 2005).

The problem facing a researcher, informed by data studies scholarship, acutely aware of the power relations in the field as well as their own position in relation to the configurations of power and knowledge (Jensen & Lauritsen, 2005: 60), and in principle sympathetic to the data activists’ normative commitments, is this: how to produce scholarship that is rigorous as well as relevant to the field, but at the same time does not under-problematise engagement? The research done in this thesis provides one possible response to this question: by maintaining a critical approach, but shifting between modes and depth of engagement. In *Why has critique run out of steam*, Bruno Latour remarks of the potential power of critique and of the role of the critic: “the critic is not the one who debunks, but the one who assembles [...] the one for whom, if something is constructed, then it means it is fragile and thus in great need of care and caution” (Latour, 2004: 246). Contending with the problematics of knowledge and power in social science – particularly the knowledge of the researcher and the relationship between the researcher and the practice – Casper Jensen and Peter Lauritsen note that “we need many more ways of linking with

practices, and not methodical, theoretical or reflexive techniques for severing the existing ones” (Jensen & Lauritsen, 2005: 72). Throughout this research, my aim has been to find a balance between knowledge production, critique and engagement – to not just “debunk”, but to also “assemble”, in a way that both produces knowledge that is relevant in the context of the data studies literature, and is also relevant for practitioners. Articles I–IV represent my humble attempts to do this. Each article maintains a commitment to problematising encountered assumptions to a more or less pronounced extent, but at the same time there is a clear shift between different modes of engagement between them.

Of the publication in this thesis, Article I is most “engaged” in the sense that it contributes knowledge to problem definitions that were encountered in the empirical field. The research took off with the aim of analysing the effectiveness of the proposed solution to the problem; the commitments encountered in the field were accepted, and their merit analysed while remaining within the defined problem space. Nevertheless, rather than simply contributing knowledge to the pre-set problem definition, such as identifying “factors” that facilitate change or “barriers” that hamper change (Zuiderent-Jerak, 2015: 8), the article ends up suggesting that the problem space itself is ill-defined. Research reported in the article, then, did not remain content with the encountered problem definitions; based on the analysis, it also led to the problematising of the encountered definitions. Articles II and III move towards more critically-oriented production of knowledge of the empirical field and analysing different aspects of what the empirical field is about. The research reported in the articles is firmly committed to problem definitions that do not emerge from the empirical field, but from data studies scholarship. Of the publications in this thesis, their mode of relating to the empirical field is the most “distanced” – however, this is not to claim that I would have been a detached outsider-observer of the data activism field during the research. Instead, the point is the position taken with respect to the production of knowledge, which is oriented at serving interests emerging from scholarly commitments rather than from the empirical field.

Article IV presents research that both engages with the empirical field, and problematises this engagement. It attempts to find a way to combine the problem definitions encountered in the field with the problem definitions emerging in critical scholarship, in effect attempting to reconfigure the encountered problem space. Therefore, during the research, we gained also a role in shaping and mobilising data activism. Our attempt at shaping data activism was obviously informed by our commitment to a critical scholarly position. However, rather than aiming to take data activism into a predetermined direction, we aimed to “open practical and analytical space for the exploration of the sociotechnical future currently in the making” (Article IV: 9). This meant that in order to produce new knowledge, we needed to “come up with ingenious solutions to the problem of how to become interesting

enough” (Jensen & Lauritsen, 2005: 72) for data activism practitioners. The approach taken in Article IV can link with practice in a way that Jensen and Lauritsen suggest: research that is “about exploring, not alone but with others, how diverse agencies could become more expressive in the invention of the future,” that is, “*exploring common futures with practices*” (Jensen & Lauritsen, 2005: 73; original emphasis). The research in Article IV can also be conceptualised as something resembling Teun Zuiderent-Jerak’s (2015) situated intervention. Situated intervention is “a scholarly approach in which intervening aims at producing sociological knowledge by situating such interventions in sociologically unpacked normative complexities” (Zuiderent-Jerak, 2015: 23). Significantly for our research, a situated intervention as discussed by Zuiderent-Jerak does not signify something that happens from the “outside” and that is separate from the practice it targets; instead, it is accepted that involvement in the field unavoidably has consequences for the researcher’s normative conceptions about the field that result from it. Intervention, then, is made in order to produce scholarly knowledge about the field under study, with awareness that this also produces normativity with respect to the field. While Zuiderent-Jerak discusses situated intervention in the context of material reconfigurations of medical practice, our modest intervention in data activism practice was more discursive than material in nature. Nevertheless, the purpose of our involvement in the practices under study was similar: to initiate a change in them, with the aim of learning something.

This alternating between modes of engagement in the articles is reminiscent of what researchers do with triangulation in different contexts. Triangulation refers to, for example, using two or more methods or theories (Jick, 1979) to interpret or examine complex empirical phenomena, or using several means to ascertain validity of research (Creswell & Miller, 2000). Triangulation, in a general sense, “may be used not only to examine the same phenomenon from multiple perspectives but also to enrich our understanding by allowing for new or deeper dimensions to emerge” (Jick, 1979: 603). In a similar sense, different modes of engagement with the empirical field can allow the enrichment of understanding and the emergence of new dimensions of the empirical field. Viewing the field from the standpoint of different commitments also allows for reflexivity not only in terms of interpreting empirical data and interpretation of the analysis process, but also in terms of the researcher’s relationship to the researched. As I have discussed, triangulating with engagement positions can lead to more comprehensive production of knowledge about the field, as well as the production of a position with respect to the normative commitments encountered in the field.

# 6 Findings

In this chapter, I discuss the findings of Articles I–IV from the perspective of the research aims and research questions of this thesis. Details beyond what is presented here can be found in the original publications. In addition, the articles’ contributions towards addressing the research questions are summarised in Table 3.

## 6.1 Article I: Consent intermediaries

In Article I we discuss personal data control services as an abstract category of intermediary services we called consent intermediaries. For an individual user, the intermediary service aims to unite the provisioning of consents under one control point, providing an access point through which individuals grant, view and withdraw permissions to collect, share, access and use data. As the frame of analysis, we consider these intermediary services from the point of view of cost–benefit evaluations on data disclosure.

In the article we first identified from literature findings on difficulties that currently hinder performing cost-benefit analysis on the collection and use of personal data. These difficulties are listed in Table 4. Many of them are related to a considerable information asymmetry: non-experts know little about collected personal data, what is done with the data, or the business operations of the data industry. Put together, they affect cost-benefit analysis by making it hard to appraise the situation and by diminishing the possibilities of making preferred decisions. We then analysed the extent to which a data control intermediary can help overcome these difficulties. We argue that there lies some potential in leveraging the intermediary position to provide various aides for making cost-benefit analyses. First, it is possible for an intermediary service to employ consent metadata across users to provide people with recommendations, predictions, ratings and similar decision aides. Second, some of the routine decisions could be automated by an intermediary service acting on behalf of the user in limited situations, such as by automating certain decision. People could, for example, choose to automatically follow recommendations made by privacy advocates. Third, the intermediary service could offer a way to leverage the data economy’s dependency on people as data

**Table 3.** The articles' contribution to the research questions of the thesis.

	<b>RQ 1a</b>	<b>RQ 1b</b>	<b>RQ 2a</b>	<b>RQ 2b</b>	<b>RQ 3a</b>	<b>RQ 3b</b>
<b>Article I</b>	Considers the ways in which an intermediary service between users and firms affects decision-making on data disclosure	-	-	-	Identifies problems that concern data arrangements where individuals control data	-
<b>Article II</b>	Outlines the forms of data agency that exemplary PDSs imagine for their users	Explores the systemic effects of data agency is imagined to have on value creation in the data economy	-	Outlines PDS developers' assumptions about users and about systemic effects of the developed technologies	-	-
<b>Article III</b>	Explores framings of resolving injustices in the data economy by construing people as agentic participants	Describes how individual data agency is imagined to lead to a competitive marketplace	Identifies alternative forms of citizen participation in the digital environment that are at odds with one another	Shows how certain forms of participation can more easily support private sector interests in data activism	Identifies forms of participation that are broader in terms of the issues that they can take into consideration	-
<b>Article IV</b>	Describes how data activism is focused on individual autonomy and empowerment	Expands on the systemic effect on the data economy imagined to be achieved by individual-level data practices	Unpacks different imaginaries underlying data activism and details tensions between them	Describes how certain imaginaries resonate with technology developers' attitudes and attachments	Problematizes the technological imaginary, outlines a socio-critical imaginary, aims to combine them in data activism	Details a situated intervention in data activism practise based on the notion of collective data governance

sources to collectively bargain or pressure for more favourable data arrangements. As Table 4 shows, some of the more practical difficulties hampering cost-benefit analysis can in principle be solved by these means. Some of the more conceptual difficulties, in contrast, arise from the privacy self-management model focusing on individuals and private cost-benefit analysis on data, and as such likely remain insuperable. The remaining difficulties remain somewhere in between; clever design could mitigate, but likely not fully resolve, them.

**Table 4.** Privacy self-management difficulties and the potential to overcome them discussed in Article I.

The difficulty	What it is about	Potential to overcome it
The timing and duration of consent	Consent is provided beforehand and indefinitely in hope of immediate benefits, while harms may develop gradually over time and are affected by future data technologies	In principle solvable; a practical problem of making it feasible to revisit decisions and revoke consent
Non-negotiability of consent	Consent is typically a “take it or leave it” choice made in terms that are dictated by the service provider	In principle solvable; a problem of negotiating power
The scale problem	There are many consent decisions to make and it is virtually infeasible to be informed in all of them	In principle solvable by making each decision easy enough
The aggregation of data	Collecting together data over individuals and contexts, and their subsequent analysis, leads to the revelation of latent data	Challenging but possible to mitigate by increasing awareness of latent data
Downstream use of data	Personal data are transferred to new parties as a result of intentional and unintentional “leaks”: e.g., data sales, data brokering, hacking, and governmental surveillance	Challenging but possible to mitigate by providing information on consented and traceable data flows
Cognitive demands	Making rational cost-benefit analysis is affected by the limitations of human decision-making: limited resources, heuristics, reasoning shortcuts, etc.	Challenging but possible to mitigate by changing the nature of decisions by, e.g., partial automation
Social norms	Consent decisions are embedded in a network of social relations that regulate behaviour and in effect restrict choices	Cannot be resolved within the individuated self-management model
The social nature of data	Personal data does not concern only one individual, and therefore privacy decisions affect others; conversely, our privacy is affected by the decisions of others	Cannot be resolved within the individuated self-management model



In Article I, the framing of the problem was a particular conception of privacy – the self-management of informational privacy based on individual cost-benefit analysis. Precisely due to this framing, the article’s findings contribute towards the broader research interest of this thesis. The difficulties we identified as hampering cost-benefit analysis, and the extent that they are amenable to circumvention by means of introducing a data control intermediary, inform us more generally of personal data as an object of value exchanges. As discussed in the article, individuals are ill-positioned to determine how such exchanges will eventually play out, and will remain ill-positioned even if they are provided an intermediary technology aiming to help in performing valuations. Personal data, in other words, are not easily amenable to individual calculations of value. The article also points out how decision-making power is transferred in various ways to the service occupying an intermediary position between people and firms. The intermediary could organise collective action and data governance, which could work favourably for individuals. Alternatively, the intermediary could affect the behaviour of its users via discreet nudges, default choices, or limitations of options, leading to the possibility of coaxing users towards specific behaviours.

## 6.2 Article II: Personal data spaces

Article II is an analysis on three exemplary data control intermediaries, called personal data spaces or PDSs, as outcomes of alternative social imaginaries of how the data economy should work. The analysis focuses on what people are imagined to do with the aid of these services, or more specifically, what kinds of agency towards data do they are imagined to afford people. The second interest is in how these forms of agency intervene in dominant ways of producing and monetising value in the data economy. The lens employed in the article was constructed based on Shoshana Zuboff’s (2015) description of surveillance capitalism, and the imagined new forms of data agency were examined in relation to surveillance capitalism’s data arrangements.

The analysis highlights four aspects of agency imagined to be made possible by PDSs: collecting data, intermediating data, controlling data analytics, and signalling subjective data. *Collecting data* happens through users’ actions of inclusion, exclusion, and moderation of data in the PDS; data may be uploaded, input, collected by sensors or devices, or transferred from elsewhere. The purpose of collecting data is not only to store, view, curate and reflect on data, but also to allow doing things with data. One such thing is *intermediating data* between primary data sources and third parties. In practice, this means providing third parties with access to data that are initially transferred into the PDS from other services. In effect, data move from one party to another via the PDS. Another one is *controlling analytics* run on data.

PDSs allow performing analytics so that “raw” data are processed within the service, and only processed data or analysis results are shared with third parties. In addition, users are imagined to *signal subjective data* to third parties. Users are imagined not only to transfer existing data into the PDS, but to create and store data on interests, preferences or intentions that are timely and relevant from the user’s point of view, and then share these data with third parties.

Based on the analysis in Article II, data agency is aimed at intervening in data economy’s dominant arrangements in different ways. Table 5 lists the key features of this imaginary of an alternative data economy.

**Table 5.** The data economy imaginary underpinning PDS services examined in Article II.

<b>What users are imagined to do</b>	<b>How it is imagined to intervene in dominant data arrangements</b>
Users control data collection	Through accumulating data in a personal repository, decisions on data collection and storage become subject to reciprocities, negotiations and feedback. Users are in control of what data are accessible, processable, shareable and available. Decision rights on data use are to remain with the data subjects.
Users supply data	Users are expected to allow third parties to access data when this provides beneficial outcomes, and have the incentive to seek new uses for their data in exchange for benefits. Through the market mechanisms, this is imagined to open up data resources to new service providers.
Users as gatekeepers	From the perspective of third parties, value production process would not begin with the extraction of data about people. Instead, it would begin with gaining access to data already stored and under the user’s control. Here, people are imagined to act as gatekeepers for third-party data use.
Users produce intermediate products	Data processing and value production performed by third parties does not necessarily happen with “raw” data, but with already processed data. Users become participants in value creation by turning data into something resembling intermediate products.
Users delimit knowledge production	By controlling analytics performed on data, users limit undesired uses of data by pre-empting knowledge production based on data by third parties. This pre-emption is another form of keeping decision rights concerning data with the user.
Users as sources of knowledge	By signalling subjective data, users become the direct sources of knowledge that service providers currently aim to produce based on data. The need to produce knowledge via data analytics is circumvented, and the quality and accuracy of services based on predictions and recommendations is improved with these data.

The imagined data agency reflects efforts to reshape the economic role of users in the data economy. Users remain the sources of personal data that keep the data economy running, but in an altered sense; users have new roles in different points of the value chain. They are suppliers of “raw” data, data-based intermediate products, or pieces of knowledge that can be thought of as the final products of data processing. An explicit aim of providing these forms of data agency is to increase the quality, accuracy and intimacy of data, in order to achieve more efficient personalisation and more accurate targeting. In this respect, the goal is to intensify datafication. While the quality or quantity of data that businesses can access is expected to increase, so is the ability of users to exercise control over the uses of data.

It is, therefore, clear that in this imaginary of the data economy, the commercial use of data is not shunned. Datafication is posited as given, and with the said interventions, datafication is imagined to lead to desirable outcomes. Commercial data use currently happens in the context of markets oriented to serve advertisers and other businesses. The alternative data economy imaginary exhibited by PDSs contains a reorientation of these markets in order to change who benefits from datafication; the data markets are to be consumer driven, and users could reap more of the produced benefits and be able to align themselves more efficiently with the market. The assumption is that new opportunities for valuable services emerge as users peruse market offerings for desirable uses for data, and exchange data for things they desire. Here, personal data effectively turn into an object of exchange. User involvement and empowerment is expected to happen through market mechanisms, which are assumed to ensure that services are designed for users to choose from. In the data economy thus envisioned, people are to be active, data-supplying and benefit-demanding, subjects and participants in value creation. For companies, this imaginary provides an alternative pathway to market success. A company could thrive by promising valuable services and analytics based on consumer-provided data.

In terms of the research objectives of this thesis, Article II identifies an alternative imaginary for the data economy, one based on a market-oriented view of data control and data agency. It provides a detailed articulation of the elements of this market agency, and how it relates to the dominant data economy imaginary.

### 6.3 Article III: Data agency at stake

Article III switches the focus from the technologies under development to considering a data activism initiative, and examines the tensions that emerge between activist and commercial interests in a situation when commercial actors are involved in data activism. The empirical context is the MyData conference. We identify collective action frames constructed by keynote speakers and the reception

of these frames by the audience. This informs us on the different ways of framing the obstacles to equal participation in the data economy, and the means of their removal. An overview of the identified collective action frames is presented in Table 6.

Based on our analysis, dominant data arrangements were framed as limiting participation of both people and firms in the data economy. For people, the primary obstacle hampering equal participation was their inability to act in relation to their personal data. For firms, the primary obstacle was access to data, which hindered commercial opportunities of everyone except those that had the capabilities to amass data in proprietary databases – namely, data economy’s dominant firms. Not surprisingly for a MyData-themed conference, a wide agreement was present regarding framing the means of resolving these problems: the development of data arrangements that provide people with data agency, or the capability for intentional action in relation to data. Developing such data arrangements was framed to be possible due to two concurrent developments. First, technological evolution was making it possible to provide individuals with data-related tools on par with those possessed by firms. Second, the evolution of formal regulation, particularly EU’s General Data Protection Regulation and its data portability rights for individuals, was providing the opportunity to access and utilise personal data held in proprietary databases. Data agency would make it possible for people to participate in data collection, data sharing and data processing, transforming personal data to serve people’s own, rather than only firms’, interests. Simultaneously, these technologies were framed as a means to redistribute data between firms, as people would redirect data to uses that they considered beneficial. This would lead to data access and related business opportunities for firms presently lacking that access.

**Table 6.** The collective action frames and their features identified in Article III.

Frame		Reception of the frame	Keynotes employing the frame
Participation enablers	Technical & legal tools	Agreed on	9
Means of participation	Agency for individuals	Agreed on	8
	Redistribution of data	Agreed on	9
Aims of participation	Beneficiaries	Contested	3
	Market symmetry	Contested	4
	Fundamental rights	Contested	2

Our analysis, therefore, identified agreement on individual data agency as being ultimately at stake. The agreed-on goal for MyData was to transform people into agentic individuals who manage their lives on- and offline and become fully-fledged participants of a datafied society. This agreement was reached without specificity on what the modes and aims of participation should be. The main contested issue concerned what participation involves. Here, data agency was framed either narrowly in terms of market choice, or more broadly in terms of data citizenship.

Framing data agency in terms of market symmetry meant narrowing participation down to market exchange. Participation signified primarily the ability to choose between alternative uses for personal data at the marketplace. People, similarly to firms, could then consider personal data as an economic asset via which to advance their own interests. The solution to data economy's woes was to provide people with the means to exchange their data for whatever suits their private interests. The promise for firms was a level playing field where access to people's data would be gained by providing them with enticing services. Here, the ability of firms to exploit data for competitive advantage would not stem from an advantageous position making efficient data extraction possible, but from success in fair competition.

Framing agency in terms of citizen participation meant transforming people not solely into market participants but into digital citizens more broadly understood, with rights, entitlements and the ability to participate in governance of data use. This frame presents a broader of participation in the economy than market participation; the economic could be considered not only as meeting market demand but, more generally, as the production of things that meet the needs of humans. Framed in this way, data activism aiming at alternative data arrangements should first consider which data arrangements allow meeting human needs and then develop data-related technologies aiming at those data arrangements. This could involve the inclusion of other kinds of value derived from data, in addition to the competitive value gained from data that others do not have.

The alternative frames of participation are particularly informative considering how the presence of firms relates to the goal-setting of data activism. Framing participation in terms of market agency was easily transformed to serve commercial data uses and allowed for the articulation of competitive benefits for the associated firms. When data agency must serve both activist and commercial interests, and market agency is readily transformed to serve commercial data uses, what is at stake risks being reduced to participation in data markets. Particular framings of data agency can thrive when data activism's agenda beyond enhancing agency remains open. In the article, we conclude that redressing injustices beyond distributive ones and deriving value from data beyond competitive advantage may hinge on data activism developing a normative agenda for what participation in a datafied society

should involve, and also on articulating citizenship-oriented data agency in more concrete terms.

Viewed in connection to the research objectives of this thesis, the frame analysis of Article III shows how individual data agency is a basis for alternative data economy imaginaries. In line with Article II, it identifies a market-focused imaginary that envisions users as participants in data markets. However, Article III also identifies an alternative, citizenship-oriented imaginary that provides alternative, broader content for data agency.

## 6.4 Article IV: The social imaginaries of data activism

Article IV continues with investigating MyData activism. Based on participation in research projects with data activists, it analytically separates two different imaginaries emerging from MyData activism. These are a technological imaginary most closely represented by technology developers, and a socio-critical imaginary that we ourselves had internalised through training in the social sciences, and that we also associated with data activism. The article has two related aims: first, to clarify the kinds of political and social alternatives that different social imaginaries ascribe to the notions underlying data activism, and second, to rework these imaginaries into a shared dialogue.

The notion of individuals having lost autonomy and being exploited by the techno-economic system that is the data economy is inherent in how MyData frames the problem at hand. As the article points out, the core ideas of the MyData vision have particular resonance with Winner's (1978) formulation of reverse adaptation, wherein the human adapts to the power of the technological system and not vice versa. Through a Winnerian lens, MyData can be viewed as being concerned with a gradual loss of control over technological arrangements. Individuals lack the power to control the system through markets or through regulations such as data protection and antitrust. What ultimately comes under threat is human autonomy; when the technological system treats humans as mere means to an end, humans are instrumentalised as sources of data rather than treated as ends as such. The technological imaginary underlying attempts to develop alternative data arrangements favours new data infrastructure as a corrective measure. Accordingly, the promoters of MyData aim at tackling reverse adaptation by suggesting that people need correctly positioned technology to be capable of self-determination. In the technological imaginary, an alternative distribution of data – or, more specifically, an alternative distribution of decision rights and capabilities related to data – generates personal and social advantages by way of economic transactions.

MyData is therefore seen as giving rise to new business models, with economically more balanced use of personal data as their driving force.

The socio-critical imaginary, in turn, questions the effectiveness of technological correction. It is informed by the critical stance characterising social scientific inquiry, drawing particularly from data studies and critical political economy to question the optimistic and future-oriented imaginary of technological advancement. In the article we outline the problematics inhering in the articulation of citizen and consumer agency in terms of individual-centric data infrastructure. Most centrally, is this not simply another iteration of Winner's reverse adaptation? Does it lead, as its advocates hope, away from reverse adaptation or does it, through expanding datafication, encouraging further reliance on data utilisation, and further opening of data to monetisation and competition, actually end up strengthening the current system? MyData's central tenets are symptomatic of a belief that individuals can control the market, if only technology is developed correctly. Here, an obvious risk of reverse adaptation lies in the belief that markets ostensibly harnessed to serve individuals would in fact allow them to control the system. A socio-critical imaginary orients us to treat the expanding commodification of personal data as a risky effort to protect human autonomy.

**Table 7.** Suggestions for combining socio-critical and technological imaginaries for future data activist work presented in Article IV.

Collective data governance	A functioning digital service environment requires governance of technical and operational rules. They could be coupled with explicit governance of data usage and exploitation, abiding to collectively agreed notions of acceptable an undesired data use
Production of data commons	Aggregate proprietary personal data and other data sources into data commons. The intended uses of data commons are intrinsically linked to the collective needs of the communities producing and governing them. Therefore, MyData can aid specific collectives in claiming personal data to benefit the community instead of only the individual. This adds a societally oriented layer to its technological infrastructure.
Learn and benefit from existing initiatives	Collective data governance, as well as data commons, could benefit from existing initiatives. These include the tradition of cooperative-based governance, as well as collaboration with social movements to demonstrate, in practice, how data can shape knowledge practices and generate advocacy and public benefits.

Article IV argues that these different imaginaries inform different kinds of engagements with production of information and knowledge. In particular, they have different relations to individual control of personal data. When viewed through the technological imaginary, MyData is an ambitious political project advancing human-

centricity. In terms of the socio-critical imaginary, it risks falling short from reaching its aims of empowerment and citizen-centricity. Well-executed MyData principles could nevertheless aid in promoting more just and sustainable data practices. To be more promising, we argue that data activism like MyData should explicitly outline intended aims for technology development, and take a normative stand on desired and undesired objectives of data usage. With this in mind, Article IV synthesises a productive relationship between the two social imaginaries, combining the infrastructure-level technological vision with explicit knowledge practices and clearly enunciated societal outcomes. As discussed in the article, towards this end, we performed a situated intervention in MyData activism by initiating discussions on “our data” with the aim of promoting collective engagement through data activism. Based on this work, the article discusses some examples of future data activist work to explore these possibilities. These are listed in Table 7.

Of the articles in this thesis, Article IV provides the broadest view into data activism. It contributes to the aims of this thesis by, first, connecting the technological imaginary animating data activism with the market-oriented imaginary identified in the previous articles; second, identifying problems inhering in the solutions developed based on this imaginary; third, developing further the notion of citizenship-based imaginary; and fourth, discussing how more sustainable data economy imaginaries can be developed both conceptually as well as in practice.



## 7 Discussion

The dominant data economy imaginary consists of notions about how firms make, and should make, use of datafication in value creation. Chapter 3 highlighted two central and interrelated aspects of the dominant imaginary. The first concerned the understanding of data as an economic resource, and how the use of data is related to competitive success in the data economy. Data are considered as a resource out there for the taking, and their exploitation for economic profit is considered as a legitimate mode of operation. Further, in order to use data for competitive advantage, a firm needs to take hold of data and prevent others from making use of them. When these notions about data are accepted, there are not many incentives for a commercial actor to forego the opportunity to extract and exploit data, as this would give edge for competitors. This is one of Jens Beckert's (2016) arguments about fictional expectations: collective imagination pushes actors under competitive pressure to accept and act according to what are presumed as normal and expected modes of action. Marion Fourcade and Kieran Healy (2017) call this the data imperative, and it incentivises embracing data economy's dominant data arrangements and modes of operation. The other aspect discussed in Chapter 3 concerned the role that these data arrangements shape for people as users, consumers, citizens and participants in the data economy. The role is that of the target of data extraction and, therefore, the source of data and the target of behavioural prediction and modification (Zuboff, 2015). The use of personal data affects things that matter in people's lives, and no doubt has also positive outcomes both individually and societally. These outcomes, whether beneficial or harmful, are however shaped by others' interests, often commercial ones, that may or may not be aligned with people's own. In short, under data economy's dominant data arrangements, people's agency towards data and their possibilities to participate in processes that determine how data are used are limited.

In order to further respond to the research questions of this thesis, this chapter discusses how, in light of the finding of Articles I–IV, the alternative data economy imaginaries developed in data activism compare to this dominant imaginary. The alternative imaginaries investigated here start from problematising some of the taken-for-granted notions about the data economy, aiming to combine notions of citizen empowerment and firms' commercial interests towards data. As the findings

discussed in Chapter 6 show, the aim is to resolve two asymmetries in the data economy. The first asymmetry is the one between people and firms or other organisations that collect and use personal data. Here, the asymmetry concerns capability of accessing and using data. Organisations are capable of making use of data to further their own ends, whereas people's capacity to do so is limited. The second asymmetry is the one between those organisations that can access data for their value creation efforts, and those that do not. This asymmetry stems from a favourable position that currently allows only some to accumulate data. Online platforms, for example, are well-positioned to accumulate data; acting as intermediaries of interactions between others, they not only make possible but also datafy those interactions.

As discussed in the articles, both asymmetries are imagined to be resolved by an alternative, in a distributive sense more symmetric data economy, consisting of a new ecosystem of services making use of personal data. Different service providers are imagined to occupy niches in the ecosystem as, for example, data producers, data users, intermediaries or infrastructure providers. Gradual expansion of the new ecosystem is imagined to lead to a data economy where individuals are a central node and control point of data flows and exchanges. Co-operation with actors that commercially utilise personal data is therefore part and parcel of the imagined alternative data economy.

The following sections are roughly organised around the themes of the research questions 1–3 of this thesis. To expand on research question 1, I begin by discussing data agency as the notion running through the findings of the articles. After that, I focus on two alternative data economy imaginaries that are both based on the notion of data agency, but nevertheless imagine different modes of societal participation implied by data agency. I refer to these two alternatives as the *market imaginary* and the *citizen imaginary*. Continuing with a further discussion on research question 2, I will outline the politics of imagination of these alternatives. The potential for success of the two alternatives, in the sense of their potential for expansion to broader agendas beyond data activism, will also be discussed. Following up with the themes of research question 3, I will finish with problematising some aspects of the developed imaginaries, and propose an outline of what may be considered, based on this research, as a more desirable data economy imaginary.

## 7.1 Data agency and the politics of imagination

Data activism examined in this thesis is in agreement on the issue at stake: data agency, understood as the individuals' capacity to act intentionally in relation to personal data and their collection and use by different actors. Broadly understood, notions about people's agency in datafied times are present in other data activism

initiatives. The concern with turning data to serve people's own ends can serve as an impetus for data activism in the first place, and this general aim underpins open data activists in Stefan Baack's (2015) research, Andrew Schrock's (2016) civic hackers, as well as different ways of using data as advocacy instrument discussed, for example, by Jonathan Gray and colleagues (2016) and Stefania Milan and Miren Gutiérrez (2018; Gutiérrez, 2018a; 2018b). In MyData activism, however, data agency can be viewed to explicitly concern participation in an economic sense, so that the aim is to improve people's condition by developing alternative data arrangements corresponding to an imaginary in which individuals are data economy's empowered actors and participants.

The pursuit of data agency in this sense implies that people's current position in relation to the techno-economic system of the data economy is seen to be subordinate. A subordinate subject position and qualities associated with it, such as passivity or helplessness, are far from proper agentic individuals who are imagined to manage life and carry the associated responsibilities (Meyer & Jepperson, 2000). Here, the subject position in the data economy is particularly a question of participation in decisions concerning data and knowledge production, and the existing, or dominant, data arrangements are seen to delimit people's possibilities of participating in these decisions. The means to alter individuals' subject position – an alternative technical and commercial service ecosystem based on the notion of individuals controlling their data – underlines that the question of agency becomes to concern the kinds of technology that are available. This suggests that certain kinds of data technology are viewed as a condition for having agency in a datafied environment. In the imagined alternative data economy, people are agentic individuals managing their lives in situations that involve personal data – in datafied times, increasingly many situation in their lives both on- and offline.

Despite the agreement on data agency being at stake, however, this research has identified two alternative modes of participation that were imagined to be the outcomes of individual agency (Article III). I argue that these modes of participation correspond to two alternative imaginaries about people's role in the data economy: the *market imaginary* and the *citizen imaginary*. The market imaginary essentially foregrounds agency as individual market choice. People are imagined as active participants in the data economy in a very specific sense; they make choices on the market about the collection, sharing, and use of their personal data. Individuals peruse market offerings and make choices in order to improve their lives and achieve better outcomes for themselves. Individuals, in other words, make personal data serve their own ends. The citizen imaginary, in contrast, frames economic and societal participation more broadly than as participation in data markets. In this imaginary, the enactment of individual control to personal data via new data technologies is viewed as a condition for realising data agency, but by itself an

inadequate means of enabling participation in the data economy. Where the market imaginary relies on market forces for data governance, the citizen imaginary views reliance on market governance as an insufficient or even dubious means to address data economy's asymmetries and to ensure favourable outcomes. In the citizen imaginary, data agency is understood as a form of civic agency, related to the capability to participate in the processes that determine how, and for what purposes, data are used. Where the market imaginary focuses on private benefits and relies on the market to provide societally desirable outcomes, the citizen imaginary foregrounds the use of data for collectively, in addition to privately, beneficial outcomes.

The division between these two alternatives is not exclusive to data activism examined in this thesis. It is aligned with a division identified by Barbara Prainsack (2019), who discusses the arguments about addressing data economy's power asymmetries as falling into two main strands: arguments for either individual or collective control. In line with the market imaginary, the individual control advocates identified by Prainsack argue that data agency and societal participation are enacted by "the implementation of ever more granular ways of informing and consenting data subjects" (Prainsack, 2019: 2). In contrast, similarly to the citizen imaginary, Prainsack's collective control advocates emphasise that "increasing individual-level control over personal data is a necessary but insufficient way to address the overarching power of multinational companies and other data capitalists" and, instead of individual control, "foreground the use of data for the public good" (Prainsack, 2019: 2).

Examining the market imaginary and the citizen imaginary *vis-à-vis* the dominant imaginary about the data economy provides a view of the data economy as a field of struggle over the collective political imagination. For actors involved in this struggle, some future states of the data economy are more advantageous than others (see Beckert, 2016), and future imaginaries are an entry point to economic power now and in the future. Aims to establish a particular version of the data economy as the future, that is, the data economy that is not just potential but that becomes the only one possible, gives rise to interest struggles over the collective imagination. This contestation can be framed in terms of Patrice Flichy's (2007) framework of the technical *imaginaire* discussed in Chapter 2. The examined data activism is based on a vision about alternative technologies producing and using personal data. This vision acts as what Flichy calls a watershed utopia, which can accommodate alternative readings of the situation. These variations of objectives are connected to choices made during the design of data technologies, so that technology development becomes an arena of contestation. A widely accepted exemplar about a material realisation of the initial vision could allow convergence towards one particular form of technology. Such an exemplar, however, is not available; no

project has, yet, developed into something that developers and end-users could converge to follow. Beyond imagining people as agentic individuals in relation to personal data, the technology vision drawing data activists and other actors together does not provide a response to how, exactly, data economy's asymmetries should be dealt with. The specifics are up for grabs and, therefore, the object of interest struggle.

In the context of data activism studied in this thesis, the politics of imagination play out in two ways. First, the imaginary of agentic individuals managing their lives is in tension with the dominant data economy imaginary. Both the market imaginary and the citizen imaginary emerge as alternatives, or at least interventions, to the dominant imaginary. Second, to the extent that market and citizen imaginaries are at odds with one another, the politics of imagination play out as contestation within the field of data activism.

## 7.2 The market imaginary

The market imaginary about people's role in the data economy essentially depicts agency as the capability of market choice. The market imaginary was discussed throughout the four articles. Articles I and II examined end-user technologies imagined to act as intermediary services via which people exercise control towards personal data. They reflected a certain control-centricity in the imagined users' role in the data economy; in essence, the imagined role is a more efficient iteration of the principle of privacy self-management (Article I). People are imagined to participate in the management of data resources and data flows, basing their decisions on the benefits and costs accruing from data use. Article II described in detail what market agency is about, its implication for value creation processes, and the intervention it makes in the dominant data arrangements. As discussed in Articles III and IV, this reliance on market forces was not a feature of service developers' imaginaries only; it prevailed also more broadly in data activism examined in this research. Article IV detailed how data activism's technological intervention was imagined to create new market arrangements to counter the reverse adaptation (Winner, 1978) that is viewed as making people serve the ends of the techno-economic system.

When the market imaginary is compared to data economy's dominant imaginary, the alternative or the intervention concerns participation in data economy's markets, the terms under which market participation happens, and who gets to decide about these things (Article II). Unsatisfactory societal developments are viewed as deficiencies in how the markets operate, and they are to be corrected by making markets operate more efficiently. The reliance on market forces as an optimal way to encourage favourable developments is also more generally embedded in dominant imaginaries about the development of the information society (Mansell, 2012). This

is, of course, also in line with markets being widely imagined as a correct way to bring order to human and social activities (Aspers, 2011).

In the context of data activism examined in this thesis, the market imaginary is premised on intervening in data economy's market arrangements by shaping people into agentic market actors. Here, reliance on market forces is present in two interrelated ways. First, individual data agency is imagined as a means for the individuals themselves to gain access to more of the proceeds of datafication. Individuals perusing the market's offerings can choose to make data available to exactly those service providers that offer the best service available. Similar rhetoric has been employed to justify extensive dataveillance practices in the dominant data economy imaginary. For example, Joseph Turow and colleagues report a retailer describing how a customer "wants to go to a retailer that understands her, is really relevant to the lifestyle she's living, and really does pay attention" (Turow et al., 2015b: 474). In the context of the dominant data economy imaginary, the corollary for a retailer facing competitive pressures is the acceleration of dataveillance, which ostensibly makes it possible to better understand the customer. In the alternative data arrangements of the market imaginary, the competitive pressure is imagined to work differently. As data are made available by customers themselves in order to receive better services, the competitive pressure would not be towards more extensive dataveillance, but rather the improvement of service promised to the customer. As discussed in Article II, the increase in the quality and intimacy of data is imagined to lead to better service via more detailed and more accurate personalisation or targeting. The corollary for the individual is the incentive to become an active participant in this process by making extensive personal data available to the provider promising the most enticing services in return for those data – after all, who would want to forego optimal service? Shoshana Zuboff's questions about who can learn from data, and who decides about this (Zuboff, 2015; also see Chapter 1) are in the market imaginary answered with "agentic individuals decide based on subjective benefits".

The second sense of reliance on market forces becomes apparent when market choices of individuals are imagined to lead to systemic effects in the data economy, as discussed in Articles II–IV. When individuals make their own choices on sharing their data with third parties promising new, beneficial services, the data economy is imagined to be opened up for competition between firms. Individuals become the conduits through which third parties can gain access to data produced about people by the primary data collectors. This makes possible the secondary use of personal data by third parties. Access to data capital would not be dependent on a favourable position allowing the extractions of data and their accumulation as proprietary resources, nor on access to the technological means to extract data. Instead, the capability to access data would be determined by competition in the provision of

enticing services for end-users. For commercial actors, the idea of rational individuals making choices on their personal data provides a potential means to compete with other firms and gain access to data resources.

Under the market imaginary, societal change is sought through the expansion of data markets, and the aim is to subject data collection to more efficient market governance. The two asymmetries to be addressed – the asymmetry between people and firms *à la* Mark Andrejevic’s “big data divide” (2014) on the one hand, and the asymmetry between data economy’s kingpins and the less powerful organisations on the other – are both imagined to be resolved by increasing individual capabilities to make choices on data use. The market competition resulting from individual choice is imagined to resolve data lock-ins and dissolve data-enabled monopolies. The new data arrangements having individual market agency as their driving force, do not only settle into the existing data markets, but rather link data agency to new business models and new means for firms to gain access to data. It is, therefore, not simply that market agency towards data is imagined to lead to empowered individuals in the data economy; rather, individual empowerment goes hand-in-hand with the advantages that the market is imagined to produce for individuals and the society at large. The market acts as the mechanism of governance that ensures more just data arrangements, as they are imagined to be the ones that will ultimately survive in market competition.

The reliance on market forces in producing favourable outcomes is based on an understanding of data’s value in terms of data being a scarce economic resource; commercial actors are imagined to benefit from data as competitive advantage that can be derived from access to data that others do not have (Cinnamon, 2017). The economic value these actors are imagined to produce with personal data is similar to value produced in the data economy currently: personalisation, profiling and targeting; optimising systems; managing and controlling things; modelling probabilities; and building stuff (Sadowski, 2019). The demand made in data industry’s marketing material cited by Jathan Sadowski (2019: 3) – “we need to understand data as an asset and turn it into value” – was repeated, in spirit if not exactly verbatim, in the empirical material analysed in Article III. The aim is to make the data economy more efficient, in the sense of producing more value, as well as more balanced, in the sense of who gets to tap into that value. In the market imaginary, personal data remains to be considered as an asset, but not only for some well-positioned firms. Through competition in service provision data can become an asset for other firms as well, and through choices on the market for people themselves. In this sense, the market imaginary does not question the data industry’s dominant economic rationale, but rather transforms it to serve the ends of data activism.

Finally, while new technologies are imagined to be instrumental for data agency, their development is not independent of the institutional context, namely data protection regulation. The development of new data arrangements, and the market arrangements associated with them, relies on regulatory opportunities, and data activists aim to ensure that those opportunities are made use of (Article III). The particular regulatory reform that data activism benefits from is the EU GDPR and its data portability rights, which are viewed as a means to gain access to personal data that has so far been held by organisations for their own purposes. In the market imaginary, formal regulation is viewed as an opportunity that makes it possible to open up data markets for competition. By relying on market forces as discussed above, data activism can leverage that competition to forward its own ends. New institutions of data protection, such as data portability, are here imagined to act as enablers that can serve data activism's ends. This enablement is thought to work both ways; the innovations made in the context of data activism are imagined to realise further effects of the instruments of data protection regulation, such as increased competition.

### 7.3 The citizen imaginary

The citizen imaginary was in this research identified in relation to the market imaginary, and it was not as extensively articulated in the empirical material. Compared to the market imaginary, the citizen imaginary extends the notion of people's agency from individual market participants towards civic agency and participation grounded in citizen rights, collective governance, and the common good. A notion about data agency oriented towards the collective is not exclusive of individual autonomy; rather, in place of an imaginary of individuals as atomistic market actors, the citizen imaginary is premised on data citizens who are self-governing but whose autonomy includes the possibility to organise collectively (see Evans, 2017). This notion of data agency was identified in Article III, and further discussed in Article IV. In practical terms, civic agency in relation to data could include democratic governance or collective negotiations over the use of personal data, and over the sharing of value produced with personal data, as well as forms of individual and collective action undertaken to reach these ends. Here, attention is not solely on who gains the benefits of personal data use, but also on processes that determine which uses are ultimately desirable, agreeable or undesirable. Potentially, data governance in this sense could apply to collective rather than individual data resources, as in the case of governance of data commons or governance based on data cooperatives (Article IV).

The commons, here, are not understood in terms of things that are freely available for all. Rather, commons are referred to in the sense in which Elinor Ostrom



(1990) discusses them: as property arrangements that allow for collective governance over norms and rules of access as well as collective enforcement of sanctions. Commons in this sense are not free for all, but are grounded on the social systems that define and regulate them (Arvidsson, 2020). In the case of collectively governed data commons based on personal data, personal data would be in some sense aggregated over individuals and transformed into a resource to be employed for the benefit of the community of people whom the data concern. This could mean a common good of a specific community, or the public good.

In this research, the market imaginary emerged as the primary imaginary, and the citizen imaginary emerged as a reaction to observed deficiencies of the market imaginary. The citizen imaginary is not necessarily a strict alternative to the market imaginary; it can rather be viewed as its expansion, so that forms of collective data governance would be applied to curb the excesses of the data markets. Nevertheless, in the empirical material, the two were clearly at odds with one another. In Article III it was discussed that the division between the market imaginary and the citizen imaginary were present already in a formative event of the MyData community. In article IV it was discussed how we, as researchers, were involved in the production and expansion of the citizen imaginary by our small intervention, situated in data activism.

One reason for the tension between the two imaginaries is that the values they embrace are contested. In this sense, the key dividing line runs between individual and collective values. The market imaginary embraces values based on individual choice, competition, effectiveness, individual benefit and property rights. It favours an understanding of data as a scarce resource, and understanding data's value to be connected to turning that resource into economic value via productive activities oriented to serve demand on the market. The citizen imaginary, in turn, embraces values based on collective governance, common or public good, and citizen rights. While the understanding of data as an economic resource underlies the citizen imaginary, it allows for accommodating a broader range of values, such as the application of data to improve human capacities, the cohesion of communities, and the non-material qualities of life (see Cinnamon, 2017). In general, collective resources such as commons can come with their distinct logic of value, owing to the meanings attached to them and the social relations that allow them to function as a resource; as Adam Arvidsson (2020: 8) formulates, they “become valuable to the extent to which they can contribute to the distinct goals and aims that are inherent in the process [...] that sustains them”. One should not expect collectively governed data resources to be different in this regard. In this sense, they should not be considered only as a form of data governance but also a way of making data valuable for those involved.

## 7.4 The success of alternative imaginaries

Two alternative imaginaries about participation in the data economy emerge from data activism, potentially in tension with one another. Can we expect one to be more successful than the other, in the sense of extending from data activism to broader agendas beyond data activism, that is, in moving from one socio-political setting to another (Jasanoff, 2015b)? The ultimate success of either imaginary is an open question, but based on this research, the market imaginary has a number of features that can make it better-positioned in this regard.

First, the market imaginary seems to better resonate with technology developers – that is, people who implement material realisations of the imaginary. The initial driving force for data activism examined in this thesis is a community of open data activists, technologists, start-up entrepreneurs and researchers. As discussed in Article IV, the technology experts tend to build upon an imaginary favouring technological correction of what are viewed as societal problems, and upon beliefs about free-spirited individuals benefiting from advances in information technology (Article IV; Barbrook & Cameron, 1996; Wyatt, 2004). Accordingly, individuals are imagined to control the market once they are provided with the technological capabilities to do so. The combination of these tendencies seems to favour the market imaginary and the development of tools and infrastructures to be used for making decisions on data, controlling data sharing, and participating in data exchanges.

Second, the regulatory opportunity supports the market imaginary. As discussed in Chapter 3 and Article I, formal privacy regulation frames data extraction as a matter of personal information concerning individuals, and decisions that individuals make on that personal information. The new data portability rights provided by Article 20 of the EU GDPR similarly focus on individuals, and particularly on their rights to access data in machine-readable format. When data activists (Article III) and like-minded service developers (Article II) leverage data portability as means towards their own ends, it is necessary for them to build on the notion of individuals as decision-makers, and to provide individuals with tools that allow them to act as conduits for data flows. The individual-centric framing of data portability seems to naturally serve technological solutions premised on a market-oriented notion of data agency. This is not to say that leveraging data portability rights would preclude framing data agency in terms citizenship and civic agency. Nevertheless, the alignment of formal regulation and the market imaginary highlights that the state can have a role in supporting a particular imaginary of an alternative data economy through regulatory efforts. For the state, formal regulation on data portability, combined with actors turning those regulation into market opportunities can be a means to address the competitive situation in the data economy.

Third, the market imaginary concurrently serves the aim of individual empowerment and the prevailing commercial interests towards personal data.

Accomplishing change in the data economy is not only a matter of developing end-user services that offer features imagined to empower individuals as actors in the data economy, and hoping that such services become significant enough to be meaningful. Such services require an environment within which they can function, and are dependent on a compatible network of regulatory frameworks, markets and other institutional arrangements, widely enough accepted data practices, and technologically and semantically interoperable data sources and endpoints, to which the new services can connect. For example, an end-user service promising to aggregate and sell its users' data cannot function if it cannot access and transfer or copy user data, nor can it work if markets for its products do not exist or do not work economically (Charitsis et al., 2018). Likewise, the service needs to offer something for those involved. In other words, it needs to be a “compromise that can be used associate multiple partners sufficiently loosely for everyone to benefit, but sufficiently rigidly [...] to function” (Flichy, 2007: 10–11). As discussed in Articles III and IV, MyData activism's chosen tactic to address this problem is to develop a gradually expanding commercial and technical ecosystem of services to which new MyData-related initiatives can attach. To function economically, the ecosystem needs to offer something for all parties involved. Based on the research in this thesis, these requirements are served by a particular understanding of data agency: the individual capacity to act in the market for data-based services. While the long-term economic and commercial viability of this imagined ecosystem of compatible services remains an open question, it nevertheless offers something for commercial technology developers and other firms: a promise of access to currently unreachable data resources. The ideas of individual data agency, the possibility to implement them by means of technology development, and the requirement to offer something for the commercial actors involved neatly come together in the market imaginary, while the citizen imaginary had fewer concrete promises to offer for service developers. Such offers could nevertheless be made. As one example, data commons can be specified and governed in a manner that delimits but does not preclude their economic exploitation (Arvidsson, 2020; Prainsack, 2019).

Fourth, the market imaginary is better aligned with the context of contemporary capitalism, where the politics of imagination are subject to the dominance and resilience of capitalism and the legitimisation and continuation of its institutions and practices (Dencik, 2018; Fisher, 2009). This resilience can be seen in the prevalence of market metaphors in the discussion of the Internet since the 1990s (Wyatt, 2004) and in the reliance on market forces in the development of the information society despite the availability of alternative imaginaries, such as those based on information commons (Mansell, 2012). The present research highlights how in the data economy, the resilience concerns some of the underpinning notions of data capitalism, including data as a resource for the production of economic value, and access to data

as a determinant of competitive success. The market imaginary exhibits the resilience of capitalism's practices and institutions in that individual agency is imagined as something that relates to existing personal data market arrangements, and participation in those markets. The role imagined for the individual, the value that is imagined to be derived from personal data, as well as the expansion of datafication and the data markets represent interventions that in a sense intervene in and reorient, instead of overhauling, replacing or fundamentally transforming, the data economy.

In light of this discussion, it is not surprising that the forms of agency imagined by the developers of new data technologies and arrangements such as personal data spaces in Article II are oriented at construing users as market participants. Identifying ways to connect with the data economy's existing arrangements, these services seek to intervene in data economy's value production processes and attach to things that produce economic value, not construct them anew from scratch.

## 7.5 Desirable data futures

The presence of both market and citizen imaginaries highlights that data agency turns out to be a malleable objective, allowing imagining alternative forms of participation in the data economy. This malleability opens up the possibility for data activism examined in this research to serve different interests while promoting data agency, highlighting the potential for divergent societal outcomes. The above discussion points out that the forms of data agency and participation based on the market imaginary have the potential to be more successful. What should we make of this potential success? As discussed in Article II, the market imaginary shapes technology development within the data activism field towards ensuring smooth data flows and developing data control interfaces and tools for end-users. What would be the outcome when such technologies are implemented and released into the wild in the context of data economy's prevailing power structures?

Relying on market governance to ensure favourable outcomes means relying on at least two assumptions. The first assumption is that once data become more broadly available, as a result of new regulatory opportunities and data activists and firms making use of these opportunities, markets start providing beneficial services for people to choose. The second assumption is that people are capable of choosing to use just these beneficial services. However, based on this research, the possibilities for making meaningful individual decisions on data is dubious. Article I pointed out some of the problems inhering in the view of personal data as a private asset to be exploited for personal benefit. In addition, to be an effective means to provide market agency, the ability to control personal data needs to be a determinant of market power. The structural incompatibility of the data industry with the empowerment of

individuals through increased control (Crain, 2018; Draper & Turow, 2019) suggests that a capability to formally control data is not necessarily a strong determinant of power in data economy's markets. As discussed in Article IV, those in less advantageous positions might be the first to end up being exposed. Those short of financial means may be willing or forced to give up their data and privacy in exchange of goods and services, and privacy might become a prerogative to which only the wealthy can aspire to. The less endowed could be the most reliant on becoming data contributors in exchange for basic services, such as internet access, housing, electricity, healthcare or insurance. Apart from financial means, the dividing line for this new personal data divide might also be, for example, financial literacy or technological savvy.

Considering the alignment of commercial interests and the market imaginary, these reservations do not necessarily matter. Whether or not people in a real sense gain agency in data economy's markets, construing them as market actors may provide firms with access to more, and more nuanced, data (Article III). Like transparency that may act as a Trojan horse via which other political goals are pursued (Levy & Johns, 2016), empowerment through individual data control may similarly advance more effective data collection and commodification. Instead of empowering people, the individual ability to control data flows may put them in a position where they end up sharing more intimate details of their personal lives, with the pretext of individuals being in charge. At the same time, this could mean more competition between firms in the data economy. The market imaginary can then serve the commercial ends of firms involved, with people's data being exploited by a new group of commercial actors.

Even this view might be an optimistic one; it deserves to be asked which commercial actors can really benefit from the economic and other opportunities created by the innovations of data activism. As pointed out in Article II, data economy's dominant actors have experimented with concepts that resemble PDS initiatives, highlighting their capability to adapt to new regulations and to occupy new positions opened up by societal developments. Article III identified the issue of beneficiaries of data activism's innovations as one of the unresolved tensions in the formative event of MyData activism. One possibility, here, would be to view any interest in individual data agency as a potential for a change for the better. The other possibility would be to rather view it as unwanted co-optation of data activist ideas by data economy's kingpins. Later interactions with MyData proponents, including discussions spurred by Article III, have indicated that this tension remains to be resolved.

These considerations highlight the market imaginary's limited potential to transform people into agentic individuals managing their lives. It is therefore tempting to look at the citizen imaginary as the alternative. Instead of individual and

market control, it is built on ideas of collective or democratic data governance and common good aims. Compared to the market imaginary, the citizen imaginary implies different orientations for technology development in data activism. The citizen imaginary, being based on notions of civic agency, could focus technology development towards equipping individuals with the tools needed for collective action and decision-making. It is not focused only on individuals as data sources and data's beneficiaries; it rather contains the idea of offering data citizens the means to bind themselves to collective undertakings that can offer better protection against data economy's exploitation than atomistic and individual-centric notions (Evans, 2017). The outcomes of personal data extraction and use can be understood as collective rather than personal issues (Article I), and the citizen imaginary better allows taking into account the interests of people, collectively understood. It is perhaps better prepared than the individual-focused market imaginary to help taking into account those situations in which people's decisions on data are fundamentally concerned with data about others (Janasik, 2019). This also means the citizen imaginary could better allow taking seriously what people themselves would consider to enable "living better with data" (Kennedy, 2018). In light of this discussion, the citizen imaginary looks like a promising starting point for developing more sustainable data economy arrangements.

Despite the above reservations on the market imaginary, something also speaks in favour of the notions of individual control. Namely, it is difficult for people to contest data economy's arrangements without having the ability to express choice on the use of their personal data to begin with. Even if data control tools do not imply free choice or ensure people's capability to intentionally participate in the data economy, they might nevertheless be a starting point for promising data arrangements. This is to say that while a desirable data economy imaginary might start from enacting individual control it does not view this as an end in itself; rather, it hinges on realising that data control technologies and market governance are not sufficient by themselves. Instead of *laissez faire* and reliance on the market governance, individual control of personal data could here be combined with explicit collective governance of the use of data for knowledge production, and the purposes the new knowledge are used. In terms of practical technology development, such combining could result in equipping individuals with institutions needed for governance based on collective action (see Evans, 2017) in the data economy.

In addition, the previous section's observations on what makes the market imaginary successful should be taken seriously. The factors that can make it successful could be learned from in order to begin constructing a desirable data economy imaginary. First, the imaginaries of technology developers will necessarily take part in and affect the formation of the data economy. Criticising their imaginaries and assumptions is important, but our experiences with data activism in

Article IV highlight that for criticism to be effective, something practical needs to be offered that steers the development of technology in a more sustainable direction. Second, the market imaginary is aligned with the prevailing commercial interests towards personal data. In other words, a successful imaginary about the data economy needs to be meaningfully situated in the current political economy of data (see Sharon, 2018); an imaginary that completely denies prevailing economic interests towards data situates itself outside of it. The third, a related but broader, point is that the dominance and resilience of capitalism and its practices suggest that it is easier to find support for ideas that can be framed in a way that resonates with the institutions of markets and market competition, and that foreground market participation.

Following this line of thinking, a successful data economy imaginary includes something that resonates with different actors having an existing stake in the situation; a data economy imaginary can extend to broader agendas if it attaches to things that currently produce economic value (Jasanoff, 2015a). To see how this can be done in practice, data arrangements based on the citizen imaginary and the enactment of civic agency in the data economy require more practical experimentation in the field of data activism.

Two rough outlines of ideas stemming from the research done for this thesis might serve as pointers towards future data activist work. First, there is the idea that individuals use data control technologies to share personal data for secondary data uses, potentially including commercial ones. These data uses are collectively governed, for instance, so that personal data are shared into data commons or some other data resource that is under collective governance (Article IV). A potential context for this idea would be data commons formed for the purpose of furthering collective interests of a social group that sustains them: for instance, a patient group sharing a condition, or a group of citizens living in the same area. Scholarly work on governance of data commons has pointed out that despite there being many proposals for data commons, there nevertheless are not many proposals on commons frameworks that would be suitable for data (Prainsack, 2019). There is a long history of employing commons as resource for market activities, and the history of appropriating commons to serve others' ends is equally long (Arvidsson, 2020). The existing commons frameworks could be learned from, and new ones developed through practical experimentation.

The second idea concerns data intermediaries, such as the PDSs of Article II. The users of a data intermediary could collectively govern the rules according to which it operates. Guided by a collective decision-making process, the intermediary could take a normative stance towards the data exchanges it facilitates; instead of acting as a hands-off and supposedly neutral facilitator, the intermediary would set binding rules that all parties connected to the intermediary need to follow, in effect

governing the outcomes of data use. Practically, the rules might resemble the terms of a license: for instance, it might be permissible to use the intermediated data only a specific purpose, for only a limited time, and only under an explicit audit mechanism. This model of operation could also constitute a form of collective bargaining, in which the intermediary service negotiates on the terms of data use, exercising power on behalf of its users over those actors it bargains with.

The fact that citizen imaginary emerged from data activism itself seems like a promising starting point for these and other experiments on more collective notions of data agency. Necessary elements for experiments on citizen agency in datafied times are therefore already present in data activism examined in this thesis.



## 8 Conclusion

Doing research on data activism has meant dealing with a continuously evolving phenomenon that is pulled into different directions by political-economic interests. As a consequence, its aims and means are constantly framed differently by different actors. This means that the research approach needs to be able to handle the messiness (Law, 2004) produced by the empirical field – otherwise, more mess will be made. My approach to make sense of the mess has been to vary the engagement position in relation to the empirical field, in effect triangulating with engagement positions to examine the empirical phenomenon from multiple perspectives and to enrich understanding of it by allowing the emergence of new features. The engagement positions taken during the course of this research have made it possible to explore different problem definitions, and to problematise and re-articulate them. Varying the engagement position has proved fruitful, as it has allowed taking new perspectives guided by the phenomenon as well as the production of a position with respect to the normative commitments encountered in the field.

The dynamics of this empirical field are an intimate part of its messiness. The dynamics mean that the variations of engagement positions are fundamentally dependent on the context of research; a different entry point might have produced different kinds of engagement, which could have affected the findings. This means that while the picture produced may be rich, even multi-dimensional, it is still a series of snapshots that are framed in a specific way. If I were to start this research now, the picture would be different; others with other entry points would produce still other pictures. This is an inherent feature of this research approach, as well as of the empirical field. Many pictures, limited as they are, nevertheless make it possible to better understand the whole. I therefore conclude with implications for future research on data activism and the data economy, and practical implications for data activism.

The expansion of the alternative imaginaries about data economy to the wider social imaginary is dependent on how they can fit in with existing norms and moral values. To be successful, the alternative should connect with things that produce economic and social value and with existing structures, whether social or material (Jasanoff, 2015b). Notions about an alternative data economy resonate well with

contemporary public discourse on technology firms' role in the society as the providers of the infrastructure of our daily lives: on the powerful position they have gained in shaping things that matter in our lives, on what this implies for democratic agency and its future, and on practical solutions such as personal everyday choices on technology use, and on antitrust regulation of the technology sector. At least anecdotally, we are witnessing an increasing social and political demand for alternatives, and increasing demand for and implementation of regulatory responses to the developing situation. This thesis has highlighted ways in which such alternatives are produced in the field of data activism.

The unifying feature of the imaginaries of data activism under study in this thesis is that increasing agency and participation is imagined as an entry point into an alternative data economy. The pursuit of data agency in the data economy, understood as the capacity to act towards the processes of data production and utilisation, is what unifies this data activism and allows its consideration as one thing, despite all the messiness. This pursuit is made possible by both technological and regulatory opportunities, viewed as enablers of shaping people as agentic actors in the data economy. Data agency is a specific entry point into imagining an alternative data economy; other entry points have been proposed elsewhere, including antitrust-motivated breakup of large technology companies, setting up new authorities for regulatory oversight, or taking stronger regulatory measures towards data as monopolistic resource. If these proposals are carried forward, current and new forms of data activism might leverage them as means to their ends.

This research has identified divisions and tensions that underlie the agreed-on objective of data agency. They are visible particularly with respect to the form of participation that data agency is imagined to imply. The main findings are outcomes of this interpretive flexibility in the aims and means of data activism, which result in an ongoing contestation over the meaning and value of what is being promoted. This flexibility supports alternative imaginaries, which can leverage the same technological and regulatory opportunities while embracing contested values and working towards divergent visions of the future data economy.

These alternative imaginaries are connected to the issue of whose aims and purposes data activism promotes. While the data activism studied here aims at people's active engagement in the data economy, this aim is not necessarily the aim of laypersons, but rather that of a community consisting of experts such as digital rights activists, technology developers and entrepreneurs. The division between the interests of laypersons and experts concerns data activism generally. In the absence of well-grounded ideas about what people consider enabling them to lead better lives with data, data activism relies upon the judgments of technical elites on what alternative data arrangements should be like (Kennedy, 2018). For example, ideas about people making use of data control interfaces in order to gain the best possible

personal outcomes from their data rely on the notion of people as willing and capable of managing data as part of their everyday lives. The techno-political goals and decisions on matters that concern the lives of people become to be defined and delimited by interests driving the expert community, and their imaginaries about how the world works and what is normal. As the research in this thesis has pointed out, when these actors include commercial firms, the interests promoted can become further delimited by not simply profitability concerns, but also specific imaginaries about the data economy.

Tensions identified in data activism might in the longer run work to undermine the broader resonance of the developed data activism imaginaries, or could tend to pull apart the movement that has emerged around the notion of people's active engagement in the data economy. As the discussion in this thesis has highlighted, the innovations of data activism could be co-opted to serve the interests of data-hungry commercial actors, including data economy's current kingpins. Other movements aiming at societal change by means of technologies and products developed in the private sector indicate one possible outcome: a division between more radical and more commercial-friendly wings (Hess, 2005). The tensions might alternatively dilute the contested aspects into something that defies a strict definition and that can be accepted by actors driven by different interests (see Flichy, 2007). All of this suggests a precarious balancing act for data activism. Being too far removed from the dominant imaginary of the data economy could lead to one obscurity, remaining decidedly outside the mainstream and having no effect in the data economy's dominant arrangements. On the other hand, being too aligned with the dominant imaginary about the data economy could lead to another obscurity, appropriation and assimilation into the dominant data arrangements, much like privacy and its protection (Coll, 2014).

The data economy imaginaries developed in MyData activism are connected to the particular institutional and regulatory context from which they emerge. Having its origin in Finland, MyData is based on assumptions about formal institutions such as data protection, and society's informal institutions such as a high level of trust among individuals, commercial actors and the state. Globally, however, the data economy's institutional conditions are different. While the Nordic imaginary about how the society functions is embedded in MyData activism, distinct imaginaries are connected to datafication elsewhere (Milan & Treré, 2019). When an initiative based on individuals controlling efficient data flows is transferred from the Nordics across contexts and jurisdictions, its dependencies with the local context can become apparent. Without the aid of regulatory instruments available in the Nordics, such as data portability, the innovations of data activism might cease working, necessitating the lobbying of regulatory reforms before embarking on technology development that can leverage regulations. If the support of specific informal institutions of the

society is removed, data activism may be seen to serve different purposes altogether (see Taylor, 2018), and in place of potential for citizen empowerment, there might instead be the potential for efficient surveillance, state control or unguarded exposure to the market. As data activism's innovations are dependent on the institutional setting, transferring them is not only the transfer of technological innovations. While this points at the need to be cautious when removing data activism's innovations from their original context, it contains also another lesson: the Nordic context may be an origin for potent ideas about alternative data arrangements. The interplay between alternative data economy imaginaries and broader, collectively held notions about institutions and shared understandings of social and economic order certainly merits further research.

To conclude, I highlight two considerations for future social-scientific scholarship on data activism that emerge from this research. The first consideration is that critical engagement can take a productive role in the context of technological activism, doing more than critique. The experience with MyData has been that activists are enthusiastic about new perspectives, but to be effective these perspectives should involve something that can involve practical activist work. This could obviously mean contributing to new technology development, but it could also mean something much more abstract, such as the articulation of complex matters in ways that help with the activists' self-understanding. When social-scientific work produces meaningful insights for activists, those insights may immediately find their ways into practical activist work. While things might start living their own lives in surprising ways, this nevertheless means that critical social-scientific work can have concrete effects on data activism practice.

This is to say that critical engagement can be constructive. The second consideration is related to why it should be. Technology developers shape things that have implications on our lives, whether we want it or not. The research in this thesis points out how imaginaries underpinning data activism can work towards different societal outcomes, and critically oriented scholarly work can pick apart and distinguish these differences. This research also highlights ways of purposefully combining elements of the different imaginaries and opening up spaces for exploring something new. The role for critical but constructive scholarship can be to identify and kindle imaginaries that are more promising in terms of leading towards a more desirable data economy, a data future that we would rather live in.

# References

- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 1–4. <https://doi.org/10.1126/science.aaa1465>
- Adams, S., Blokker, P., Doyle, N.J., Krummel, J.W.M., & Smith, J.C.A. (2015). Social imaginaries in debate. *Social Imaginaries*, 1(1), 15–52. <https://doi.org/10.5840/si2015112>
- Anderson, B. (1991). *Imagined Communities. Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, June 23. [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory)
- Anderson, R. (1994). Representations and requirements: The value of ethnography in system design. *Human–Computer Interaction*, 9(3), 151–182. [https://doi.org/10.1207/s15327051hci0902\\_1](https://doi.org/10.1207/s15327051hci0902_1)
- Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, 8, 1673–1689. <https://ijoc.org/index.php/ijoc/article/view/2161>
- Andrejevic, M. (2016). Theorizing drones and droning theory. In: Završnik, A. (Ed.) *Drones and Unmanned Aerial Systems. Legal and Social Implications for Security and Surveillance*, 21–44. Cham: Springer. <https://doi.org/10.1007/978-3-319-23760-2>
- Andrejevic, M. (2017). To preempt a thief. *International Journal of Communication*, 11, 879–896. <https://ijoc.org/index.php/ijoc/article/view/6308>
- Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society*, 12(2), 185–196. <https://doi.org/10.24908/ss.v12i2.5242>
- Arrieta Ibarra, I., Goff, L., Hernández, D.J., Lanier, J., & Weyl, E.G. (2018). Should we treat data as labor? Moving beyond “free.” *American Economic Association Papers & Proceedings*, 108, 38–42. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3093683](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3093683)
- Arvidsson, A. (2020). Capitalism and the commons. *Theory, Culture & Society*, 37(2), 3–30. <https://doi.org/10.1177/0263276419868838>
- Aspers, P. (2011). *Markets*. Cambridge, UK: Polity Press.
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 1–11. <https://doi.org/10.1177/2053951715594634>
- Baack, S. (2018a). Civic Tech at mySociety: How the Imagined Affordances of Data Shape Data Activism. *Krisis*, 1, 44–56. <https://krisis.eu/civic-tech-at-mysociety-how-the-imagined-affordances-of-data-shape-data-activism/>
- Baack, S. (2018b). *Knowing What Counts. How Journalists and Civic Technologists Use and Imagine Data*. Groningen: University of Groningen. <http://hdl.handle.net/11370/4c94668a-c25c-43cb-9b36-5d54e3ff3c2e>
- Barbrook, R., & Cameron, A. (1996). The Californian ideology. *Science as Culture*, 6(1), 44–72. <https://doi.org/10.1080/09505439609526455>
- Barnes, S. (2006). A privacy paradox: Social networking in the United States. *First Monday*, 11(9). <http://firstmonday.org/article/view/1394/1312>

- Barta, K., & Neff, G. (2016). Technologies for sharing: Lessons from Quantified Self about the political economy of platforms. *Information, Communication & Society*, 19(4), 518–531. <https://doi.org/10.1080/1369118X.2015.1118520>
- Baruh, L., & Popescu, M. (2017). Big data analytics and the limits of privacy self-management. *New Media & Society*, 19(4), 579–596. <https://doi.org/10.1177/1461444815614001>
- Bates, J. (2013). The domestication of open government data advocacy in the United Kingdom: A neo-Gramscian analysis. *Policy and Internet*, 5(1), 118–137. <https://doi.org/10.1002/poi3.25>
- Beckert, J. (2016). *Imagined Futures. Fictional Expectations and Capitalist Dynamics*. Cambridge, MA: MIT Press.
- Beer, D. (2018). Envisioning the power of data analytics. *Information, Communication & Society*, 21(3), 465–479. <https://doi.org/10.1080/1369118X.2017.1289232>
- Belli, L., Schwartz, M., & Louzada, L. (2017). Selling your soul while negotiating the conditions: From notice and consent to data control by design. *Health Technology*, 7(4), 453–467. <https://doi.org/10.1007/s12553-017-0185-3>
- Bellotti, V., & Sellen, A. (1993). Design for privacy in ubiquitous computing environments. In: *Proceedings of the Third European Conference on Computer-Supported Cooperative Work*, 77–92. Dordrecht: Springer. <https://dl.acm.org/citation.cfm?id=1241940>
- Bellovin, S.M., Hutchins, R.M., Jebara, T., & Zimmeck, S. (2013). When enough is enough: Location tracking, mosaic theory, and machine learning. *NYU Journal of Law & Liberty*, 8, 555–628. <https://www.ssrn.com/abstract=2320019>
- Benford, R.D., & Snow, D.A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, 26(1), 611–639. <https://doi.org/10.1146/annurev.soc.26.1.611>
- Beraldo, D., & Milan, S. (2019). From data politics to the contentious politics of data. *Big Data & Society*, 6(2), 1–11. <https://doi.org/10.1177/2053951719885967>
- Berry, D. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, 12, 1–22. <https://culturemachine.net/wp-content/uploads/2019/01/10-Computational-Turn-440-893-1-PB.pdf>
- Beyer, J. (2014). The emergence of a freedom of information movement: anonymous, Wikileaks, the pirate party, and Iceland. *Journal of Computer-Mediated Communication*, 19(2), 141–154. <https://doi.org/10.1111/jcc4.12050>
- Biggs, J. (2018). #deletefacebook. *Techcrunch*, March 19. <https://techcrunch.com/2018/03/19/deletefacebook/>
- Bohman, J. (2016). Critical theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University. <https://plato.stanford.edu/entries/critical-theory/>
- Boston Consulting Group (2012). The value of our digital identity. *Liberty Global Policy Series*. <http://www.lgi.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf>
- Bowker, G.C., & Star, S.L. (1999). *Sorting Things Out. Classification and its Consequences*. Cambridge, MA: MIT Press.
- boyd, danah, & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Braudel, F. (1992). *Civilization and Capitalism, 15th–18th Century, Vol. III: The Perspective of the World*. Berkeley, CA: University of California Press.
- Breindl, Y. (2013). Assessing success in internet campaigning: The case of digital rights advocacy in the European Union. *Information, Communication & Society*, 16(9), 1419–1440. <https://doi.org/10.1080/1369118X.2012.707673>
- Brooker, K. (2018). “I was devastated”: Tim Berners-Lee, the man who created the world wide web, has some regrets. *Vanity Fair*, July 1. <https://www.vanityfair.com/news/2018/07/the-man-who-created-the-world-wide-web-has-some-regrets>
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 14(7), 1164–1180. <https://doi.org/10.1177/1461444812440159>

- Caplan, R., & boyd, d. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society*, 5(2), 1–12. <https://doi.org/10.1177/2053951718757253>
- Callon, M. (1998). Introduction: The embeddedness of economic markets in economics. *The Sociological Review*, 46(S1), 1–57. <https://doi.org/10.1111/j.1467-954X.1998.tb03468.x>
- Castells, M. (1996). *The Rise of the Network Society. Vol. 1 of the Information Age: Economy, Society, and Culture*. Cambridge, UK: Blackwell Publishing.
- Castells, M. (2007). Communication, power and counter-power in the network society. *International Journal of Communication*, 1, 238–266. <https://ijoc.org/index.php/ijoc/article/view/46>
- Castoriadis, C. (1987). *The Imaginary Institution of Society*. Cambridge, MA: MIT Press.
- Cattaneo, G., Micheletti, G., Osimo, D., & Jakimovicz, K. (2018). *How the Power of Data Will Drive EU Economy. First Report on Policy Conclusions*. Brussels: European Commission. [http://datalandscape.eu/sites/default/files/report/EDM\\_D2.2\\_First\\_Report\\_on\\_Policy\\_Conclusions\\_20.04.2018.pdf](http://datalandscape.eu/sites/default/files/report/EDM_D2.2_First_Report_on_Policy_Conclusions_20.04.2018.pdf)
- Charitsis, V., Zwick, D., & Bradshaw, A. (2018). Creating worlds that create audiences: Theorising personal data markets in the age of communicative capitalism. *tripleC: Communication, Capitalism & Critique*, 16(2), 820–834.
- Cheney-Lippold, J. (2011). A new algorithmic identity. *Soft biopolitics and the modulation of control. Theory, Culture & Society*, 28(6), 164–181. <https://doi.org/10.1177/0263276411424420>
- Cinnamon, J. (2017). Social injustice in surveillance capitalism. *Surveillance & Society*, 15(5), 609–625. <https://doi.org/10.24908/ss.v15i5.6433>
- Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, 31(5), 498–512. <https://doi.org/10.1145/42411.42413>
- Clarke, R. (2003). *Dataveillance – 15 years on*. <http://www.rogerclarke.com/DV/DVNZ03.html>
- Coll, S. (2014). Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. *Information, Communication & Society*, 17(10), 1250–1263. <https://doi.org/10.1080/1369118X.2014.918636>
- Couldry, N., & Mejias, U.A. (2019). Data colonialism: Rethinking big data’s relation to the contemporary subject. *Television and New Media*, 20(4), 336–349. <https://doi.org/10.1177/1527476418796632>
- Couldry, N., & Yu, J. (2018). Deconstructing datafication’s brave new world. *New Media & Society*, 20(12), 4473–4491. <https://doi.org/10.1177/1461444818775968>
- Crabtree, A., Lodge, T., Colley, J., & Greenhalgh, C. (2016). Enabling the new economic actor: Data protection, the digital economy, and the Databox. *Personal and Ubiquitous Computing*, 20(6), 947–957. <https://doi.org/10.1007/s00779-016-0939-3>
- Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, 20(1), 88–104. <https://doi.org/10.1177/1461444816657096>
- Creswell, J. W., & Miller, D. L. (2000). Determining Validity in qualitative Inquiry. *Theory into Practice*, 39(3), 124–130. [https://doi.org/10.1207/s15430421tip3903\\_2](https://doi.org/10.1207/s15430421tip3903_2)
- Custers, B. (2016). Click here to consent forever: Expiry dates for informed consent. *Big Data & Society*, 3(1), 1–6. <https://doi.org/10.1177/2053951715624935>
- Dalton, C., Taylor, L., & Thatcher, J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society*, 3(1), 1–9. <https://doi.org/10.1177/2053951716648346>
- Dalton, C., & Thatcher, J. (2014). What does a critical data studies look like, and why do we care? *Society and Space*, 12 May. <https://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/>
- de Montjoye, Y.-A., Shmueli, E., Wang, S.S., & Pentland, A. (2014). openPDS: Protecting the privacy of metadata through SafeAnswers. *PloS One*, 9(7), 1–9. <https://doi.org/10.1371/journal.pone.0098790>
- Degli Esposti, S. (2014). When big data meets dataveillance: The hidden side of analytics. *Surveillance and Society*, 12(2), 209–225. <https://doi.org/10.24908/ss.v12i2.5113>

- Dencik, L. (2018). Surveillance realism and the politics of imagination: Is there no alternative? *Krisis*, 1, 31–43. <https://krisis.eu/surveillance-realism-and-the-politics-of-imagination-is-there-no-alternative/>
- Dencik, L., & Cable, J. (2017). The advent of surveillance realism: Public opinion and activist responses to the Snowden leaks. *International Journal of Communication*, 11, 763–781. <https://ijoc.org/index.php/ijoc/article/view/5524>
- Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*, 3(2), 1–12. <https://doi.org/10.1177/2053951716679678>
- DiMaggio, P.J., & Powell, W.W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://www.jstor.org/stable/2095101>
- Dommeyer, C., & Gross, B. (2003). What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing*, 17(2), 34–51. <https://doi.org/10.1002/dir.10053>
- Domo, (2019). *Data Never Sleeps 7.0*. <https://www.domo.com/learn/data-never-sleeps-7>
- Draper, N.A. (2017). From privacy pragmatist to privacy resigned: Challenging narratives of rational choice in digital privacy debates. *Policy & Internet*, 9(2), 232–251. <https://doi.org/10.1002/poi3.142>
- Draper, N.A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*. <https://doi.org/10.1177/1461444819833331>
- Edge. (2012). *Reinventing Society in the Wake of Big Data: A Conversation with Alex (Sandy) Pentland*, August 30. <http://www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data>
- Elder-Vass, D. (2016). *Profit and Gift in the Digital Economy*. Cambridge, UK: Cambridge University Press.
- Elder-Vass, D. (2018). Lifeworld and systems in the digital economy. *European Journal of Social Theory*, 21(2), 227–244. <https://doi.org/10.1177/1368431017709703>
- Elmer, G. (2004). *Profiling Machines: Mapping the Personal Information Economy*. Cambridge, MA: MIT Press.
- Engels, B. (2016). Data portability among online platforms. *Internet Policy Review*, 5(2). <https://doi.org/10.14763/2016.2.408>
- European Commission. (2016). *An Emerging Offer of ‘Personal Information Management Services’ – Current State of Service Offers and Challenges*. Brussels: European Commission. [http://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=40118](http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=40118).
- Evans, B.J. (2017). Power to the people: Data citizens in the age of precision medicine. *Vanderbilt Journal of Entertainment and Technology Law*, 19(2), 243–265. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5673282/>
- Feenberg, A. (1999). *Questioning Technology*. London: Routledge.
- Felt, U. (2015). Keeping technologies out: Sociotechnical imaginaries and the formation of Austria’s techno political identity. In: Jasanoff, S. & Kim, S-H. (Eds.) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, 103–125. Chicago, IL: University of Chicago Press.
- Fisher, M. (2009). *Capitalist realism: Is there no alternative?* Hants, UK: Zero Books.
- Flender, C., & Müller, G. (2012). Type indeterminacy in privacy decisions: The privacy paradox revisited. In: Busemeyer J.R., Dubois F., Lambert-Mogiliansky A., Melucci M. (Eds.) *Quantum Interaction. QI 2012. Lecture Notes in Computer Science*, 7620, 148–159. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-642-35659-9\\_14](https://doi.org/10.1007/978-3-642-35659-9_14)
- Flichy, P. (2007). *The Internet Imaginaire*. Cambridge, MA: MIT Press.
- Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A*, 374, 1–5. <http://dx.doi.org/10.1098/rsta.2016.0360>



- Flyverbom, M., Delbert, R., & Matten, D. (2019). The governance of digital technology, Big Data, and the internet: New roles and responsibilities for business. *Business & Society*, 58(1), 3–19. <https://doi.org/10.1177/0007650317727540>
- Flyverbom, M., Madsen, A.K., & Rasche, A. (2017). Big Data as governmentality in international development: Digital traces, algorithms, and altered visibilities. *The Information Society*, 33(1), 35–42. <https://doi.org/10.1080/01972243.2016.1248611>
- Foster, J.B., & McChesney, R.W. (2004). Surveillance capitalism: Monopoly-finance capital, the military-industrial complex, and the digital age. *Monthly Review*, 66(3). <https://monthlyreview.org/2014/07/01/surveillance-capitalism/?v=7516fd43adaa>
- Fourcade, M., & Healy, K. (2017). Seeing like a market. *Socio-Economic Review*, 15(1), 9–29. <https://doi.org/10.1093/ser/mww033>
- Fraser, N. (2008). Abnormal justice. *Critical Inquiry*, 34(3), 393–422. <https://doi.org/10.1086/589478>
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iView*, June. <https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
- Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), 1239–1249. <https://doi.org/10.1016/j.respol.2014.03.006>
- Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15(8), 1348–1365. <https://doi.org/10.1177/1461444812472322>
- Gillespie, T. (2010). The politics of “platforms.” *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Gilson, L., & Goldberg, C. (2015). Editor’s comment: So, what is a conceptual paper? *Group and Organization Management*, 40(2), 127–130. <https://doi.org/10.1177/1059601115576425>
- Gitelman, L. & Jackson, V. (2013). Introduction. In: Gitelman, L. (Ed.) “Raw data” is an oxymoron, 1–14. Cambridge, MA: MIT Press.
- Gray, J., Lämmerhirt, D., & Bounegru, L. (2016). *Changing What counts: How Can Citizen-generated and Civil Society Data Be Used as an Advocacy Tool to Change Official Data Collection?* <http://dx.doi.org/10.2139/ssrn.2742871>
- Gutiérrez, M. (2018a). *Data Activism and Social Change*. Cham: Palgrave Macmillan.
- Gutiérrez, M. (2018b). Data activism in light of the public sphere. *Krisis*, 1, 57–71. <https://krisis.eu/data-activism-in-light-of-the-public-sphere/>
- Gutiérrez, M., & Milan, S. (2019). Playing with data and its consequences. *First Monday*, 24(1). <https://firstmonday.org/ojs/index.php/fm/article/view/9554/7716>
- Hagel, J., & Rayport, J., (1997). The coming battle for consumer information. *Harvard Business Review*, 75, 53–65. <https://hbr.org/1997/01/the-coming-battle-for-customer-information>
- Haggerty, K.D., & Ericson, R.V. (2000). The surveillant assemblage. *British Journal of Sociology*, 51(4), 605–622. <https://doi.org/10.1080/00071310020015280>
- Hargittai, E., & Marwick, A. (2016). “What can I really do?” Explaining the privacy paradox with online apathy. *International Journal of Communication*, 10, 3737–3757. <https://ijoc.org/index.php/ijoc/article/view/4655>
- Hass, J.K. (2007). *Economic Sociology: An Introduction*. London: Routledge.
- Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media + Society*, 1(2), 1–11. <https://doi.org/10.1177/2056305115603080>
- Hepp, A. (2016). Pioneer communities: Collective actors in deep mediatization. *Media, Culture & Society*, 38(6), 918–933. <https://doi.org/10.1177/0163443716664484>
- Herschel, R., & Miori, V.M. (2017). Ethics & Big Data. *Technology in Society*, 49, 31–36. <https://doi.org/10.1016/j.techsoc.2017.03.003>
- Hess, D.J. (2005). Technology- and product-oriented movements: Approximating social movement studies and science and technology studies. *Science, Technology, & Human Values*, 30(4), 515–535. <https://doi.org/10.1177/0162243905276499>
- Hess, D.J. 2014. Publics as threats? Integrating science and technology studies and social movement studies. *Science as Culture*, 24(1), 69–82. <https://doi.org/10.1080/09505431.2014.986319>

- Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance society. *International Journal of Communication*, 11, 731–739. <http://ijoc.org/index.php/ijoc/article/view/5521>
- Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2019). *Digital Citizenship in a Datafied Society*. Cambridge, UK: Polity Press.
- Hoffman, P.H., Lutz, C., & Ranzini, G. (2016). Privacy cynicism: A new approach to the privacy para- dox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 10(4), art. 7. <http://dx.doi.org/10.5817/CP2016-4-7>
- Hoofnagle, C.J., & Urban, J.M. (2014). Alan Westin’s privacy homo economicus. *Wake Forest Law Review* 261, 261–317. <https://www.ssrn.com/abstract=2434800>
- Hughes, J., King, V., Rodden, T., & Andersen, H. (1994) Moving out from the control room: Ethnography in system design. In: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 429–439. New York, NY: ACM. <https://doi.org/10.1145/192844.193065>
- Imma, R. (2016). Towards personal data services: a view on some enabling factors. *International Journal of Electronic Governance*, 8(1), 58–73. <https://doi.org/10.1504/IJEG.2016.076690>
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), 1–7. <https://doi.org/10.1177/2053951716674238>
- Janasik, N. (2019). Reframing autonomy: My data, our data, and the question of human dignity. In: Toivonen, M. & Saari, E. (Eds.) *Human-Centered Digitalisation and Services*, 245–258. Singapore: Springer.
- Jasanoff, S. (2015a). Future imperfect: Science, technology and the imaginations of modernity. In: Jasanoff, S., & Kim, S-H. (Eds.) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, 1–33. Chicago, IL: University of Chicago Press.
- Jasanoff, S. (2015b). Imagined and invented worlds. In: Jasanoff, S. & Kim, S-H. (Eds.) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, 321–342. Chicago, IL: University of Chicago Press.
- Jasanoff, S., & Kim, S-H. (2009). Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva*, 47(2), 119–146. <https://doi.org/10.1007/s11024-009-9124-4>
- Jensen, C., & Lauritsen, P. (2005). Qualitative research as partial connection: Bypassing the power–knowledge nexus. *Qualitative Research*, 5(1), 59–77. <https://doi.org/10.1177/1468794105048652>
- Johnson, J.A. (2014). From open data to information justice. *Ethics and Information Technology*, 16(4), 263–274. <https://doi.org/10.1007/s10676-014-9351-8>
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), 602–611. <http://www.jstor.org/stable/2392366>
- Kannengießer, S. (2019). Reflecting and acting on datafication – CryptoParties as an example of re- active data activism. *Convergence*, 1–14. <https://doi.org/10.1177/1354856519893357>
- Kaun, A., & Uldam, J. (2017). Digital activism: After the hype. *New Media & Society* 20(6), 2099–2106. <https://doi.org/10.1177/1461444817731924>
- Kelty, C.M. (2008). *Two Bits: The Cultural Significance of Free Software*. Durham: Duke University Press.
- Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis*, 1, 18–30. <https://krisis.eu/living-with-data/>
- Kennedy, H., Elgesem, D., & Miguel, C. (2015). On fairness: User perspectives on social media data mining. *Convergence*, 23(3), 270–288. <https://doi.org/10.1177/1354856515592507>
- Kim, S-H. (2015). Social movements and contested sociotechnical imaginaries in South Korea. In: Jasanoff, S., & Kim, S-H. (Eds.) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, 152–173. Chicago, IL: University of Chicago Press.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers and Security*, 64(January), 122–134. <https://doi.org/10.1016/j.cose.2015.07.002>
- Kranzberg, M. (1986). Technology and history: “Kranzberg's laws”. *Technology and Culture*, 27(3), 544–560. <https://doi.org/10.2307/3105385>
- Kuittinen, O., & Ruckenstein, M. (2014). Vieraskynä: Ihmisillä on oikeus hallita digitaalista jalanjälkeään. *Helsingin Sanomat*, June 13. <http://www.hs.fi/paakirjoitukset/a1402548692688>
- Lanier, J. (2013). *Who owns the Future?* London: Penguin Books.
- Lanier, J. & Weyl, E.G. (2018). A blueprint for a better digital society. *Harvard Business Review*, 26. <https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society>
- Larkin, B. (2013). The politics and poetics of infrastructure. *Annual Review of Anthropology*, 42, 327–343. <https://doi.org/10.1146/annurev-anthro-092412-155522>
- Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, 30(2), 225–248. <https://doi.org/10.1086/421123>
- Laudon, K. (1996). Markets and privacy. *Communications of the ACM*, 39(2), 92–104. <https://doi.org/10.1145/234215.234476>
- Law, J. (2004). *After Method: Mess in Social Science Research*. London: Routledge.
- Levy, K.E.C., & Johns, D.M. (2016). When open data is a Trojan Horse: The weaponization of transparency in science and governance. *Big Data & Society*, 3(1), 1–6. <https://doi.org/10.1177/2053951715621568>
- Levy, P. (2019). The prophet of Silicon Valley. *Mother Jones*, November 22. <https://www.motherjones.com/politics/2019/11/andrew-yang/>
- Lyon, D. (2007). *Surveillance Studies: An overview*. Cambridge, UK: Polity Press.
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 1–13. <https://doi.org/10.1177/2053951714541861>
- MacKenzie, D. (2006). *An Engine, Not a Camera. How Financial Models Shape Markets*. Cambridge, MA: MIT Press.
- Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, 15(5), 769–787. <https://doi.org/10.1080/1369118X.2012.676056>
- Mager, A. (2017). Search engine imaginary: Visions and values in the co-production of search technology and Europe. *Social Studies of Science*, 47(2), 240–262. <https://doi.org/10.1177/0306312716671433>
- Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, 32(3), 192–199. <https://doi.org/10.1080/01972243.2016.1153010>
- Mansell, R. (2012). *Imagining the Internet. Communication, Innovation, and Governance*. Oxford: Oxford University Press.
- Manson, N.C., & O’Neill, O. (2007). *Rethinking Informed Consent in Bioethics*. Cambridge, UK: Cambridge University Press.
- Marr, B. (2018). How much data do we create every day? The mind-blowing stats everyone should read. *Forbes*, May 21. <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#30cac87a60ba>
- Marwick, A., & Hargittai, E. (2018). Nothing to hide, nothing to lose? Incentives and disincentives to sharing information with institutions online. *Information, Communication & Society*. <https://doi.org/10.1080/1369118X.2018.1450432>
- Maurer, B. (2015). Principles of descent and alliance for big data. In: Boellstorff, T. & Maurer, B. (Eds.) *Data, Now Bigger and Better!*, 67–86. Chicago: Prickly Paradigm Press.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.
- McDonald, A.M., & Cranor, L.F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4, 543–568. <https://heinonline.org/HOL/P?h=hein.journals/isjlpoc4&i=563>

- McNeil, M., Arribas-Ayllon, M., Haran, J., Mackenzie, A., & Tutton, R. (2017). Conceptualizing imaginaries of science, technology and society. In: Felt, U., Fouche, R., Miller, C.A., & Smith-Doerr, L. (Eds.) *The Handbook of Science and Technology Studies*, 4th edition, 435-464. Cambridge, MA: MIT Press.
- Meyer, J. W., & Jepperson, R. L. (2000). The actors of modern society: The cultural construction of social agency. *Sociological Theory*, 18(1), 100–120. <https://doi.org/10.1196/annals.1352.037>
- Milan, S., & Gutiérrez, M. (2015). Citizens' media meets Big Data: The emergence of data activism. *Mediaciones*, 14, 120–133. <https://hdl.handle.net/11245/1.520998>
- Milan, S., & Gutiérrez, M. (2018). Technopolitics in the age of Big Data. In: Caballero, F.S., & Gravante, T. (Eds.) *Networks, Movements & Technopolitics in Latin America: Critical Analysis and Current Challenges*, 95–109. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-65560-4>
- Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond data universalism. *Television and New Media*, 20(4), 319–335. <https://doi.org/10.1177/1527476419837739>
- Milan, S., & van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2), 57–74. <https://doi.org/10.14361/dcs-2016-0205>
- Molla, R. & Stewart, E. (2019). 2020 Democrats on who controls your data – and who's at fault when it's mishandled. *Vox*, December 3. <https://www.vox.com/policy-and-politics/2019/12/3/20965463/tech-2020-candidate-policies-online-data-equifax>
- Monahan, T. (2008). Surveillance and inequality. *Surveillance & Society*, 5(3), 217–226. <https://doi.org/10.24908/ss.v5i3.3421>
- Monahan, T. (2016). Built to lie: Investigating technologies of deception, surveillance, and control. *The Information Society*, 32(4), 229–240. <https://doi.org/10.1080/01972243.2016.1177765>
- MyData Global (2019). *MyData Global website*. <https://mydata.org>
- Nafus, D. & Sherman, J. (2014). This one does not go up to 11: The Quantified Self movement as an alternative big data practice. *International Journal of Communication*, 8, 1784–1794. <https://ijoc.org/index.php/ijoc/article/view/2170>
- Nelimarkka, M., Kuikkaniemi, K., Salovaara, A., & Jacucci, G. (2016). Live participation: Augmenting events with audience-performer interaction systems. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 509–520. New York: ACM. <https://doi.org/10.1145/2901790.2901862>
- Noam, E.M. (1995). Privacy in telecommunications: Markets, rights, and regulations. Part III: Markets in privacy. *New Telecom Quarterly*, 4Q95, 51–60. [http://www.tfi.com/pubs/ntq/articles/view/95Q4\\_A9.pdf](http://www.tfi.com/pubs/ntq/articles/view/95Q4_A9.pdf)
- Norberg, P., Horne, D.R., & Horne, D.A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1), 100–127. <https://doi.org/10.1111/j.1745-6606.2006.00070.x>
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Park, Y. (2013). Digital literacy and privacy behavior online. *Communication Research*, 40(2), 215–236. <https://doi.org/10.1177/0093650211418338>
- Pasquale, F. (2017). Two narratives of platform capitalism. *Yale Law & Policy Review*, 35(1), 309–319. <https://ylpr.yale.edu/two-narratives-platform-capitalism>
- Pendergrast, K. (2019). The next big cheap. *Real Life*, November 25. <https://reallifemag.com/the-next-big-cheap/>
- Pentland, A. (2009). Reality mining of mobile communications: Toward a new deal on data. In S. Dutta & I. Mia (Eds.) *The Global Information Technology Report 2008–2009. Mobility in a Networked World*, 75–80. World Economic Forum.
- Perrin, A. (2018). Americans are changing their relationship with Facebook. *Pew Research Center Fact Tank*, September 5. <https://www.pewresearch.org/fact-tank/2018/09/05/americans-are-changing-their-relationship-with-facebook/>

- Peters, B. (2016). Digital. In: Peters, B. (Ed.) *Digital Keywords: A Vocabulary of Information Society and Culture*, 93–108. Princeton, NJ: Princeton University Press.
- Piketty, T. (2014). *Capital in the Twenty-First Century*, Cambridge, MA: The Belknap Press of Harvard University Press.
- Pitkänen, O. (2014). Sinun tietosi eivät ole sinun: Rekisteröidyn oikeus hyödyntää omia henkilötietojaan. *Oikeus*, 43, 202–214. <https://www.edilex.fi/oikeus/13599>
- Poikola, A., Kuikkaniemi, K., & Honko, H. (2015). *MyData – A Nordic Model for Human-Centered Personal Data Management and Processing*. Helsinki: Finnish Ministry of Transport and Communications. <http://urn.fi/URN:ISBN:978-952-243-455-5>
- Poikola, A., Kuikkaniemi, K., & Kuittinen, O. (2014). *My Data – Johdatus Ihmiskeiseen Henkilötiedon Hyödyntämiseen*. Helsinki: Finnish Ministry of Transport and Communications. <http://urn.fi/URN:ISBN:978-952-243-418-0>
- Pollach, I. (2005). A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. *Journal of Business Ethics*, 62(3), 221–235. <https://doi.org/10.1007/s10551-005-7898-3>
- Portwood-Stacer, L. (2013). Media refusal and conspicuous non-consumption: the performative and political dimensions of Facebook abstention. *New Media & Society* 15(7), 1041–1057. <https://doi.org/10.1177/1461444812465139>
- Postigo, H. (2012). Cultural production and the digital rights movement: Framing the right to participate in culture. *Information, Communication & Society*, 15(8), 1165–1185. <https://doi.org/10.1080/1369118X.2011.568509>
- Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*, 6(1), 1–15. <https://doi.org/10.1177/2053951719829773>
- Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication*, 8, 1690–1709. <https://ijoc.org/index.php/ijoc/article/view/2169>
- Pybus, J., Coté, M., & Blanke, T. (2015). Hacking the social life of Big Data. *Big Data & Society*, 2(2), 1–10. <https://doi.org/10.1177/2053951715616649>
- Raley, R. (2013). Dataveillance and counterveillance. In: Gitelman, L. (Ed.) *“Raw Data” is an Oxymoron*, 121–146. Cambridge, MA: MIT Press.
- Reiss, J., & Sprenger, J. (2017). Scientific objectivity. In: Zalta, E. (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Stanford University. <https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/>
- Richards, N.M., & King, J.H. (2014). Big Data Ethics. *Wake Forest Law Review*, 29, 393–432. <https://ssrn.com/abstract=2384174>
- Richardson, M. (2018). Drone capitalism. *Transformations*, 31, 79–98. [http://www.transformationsjournal.org/wp-content/uploads/2018/06/Trans31\\_05\\_richardson.pdf](http://www.transformationsjournal.org/wp-content/uploads/2018/06/Trans31_05_richardson.pdf)
- Ricœur, P. (1986). *Lectures on Ideology and Utopia*. Taylor, G. (Ed.). Chicago, IL: Chicago University Press.
- Rieder, B., & Sire, G. (2014). Conflicts of interest and incentives to bias: A microeconomic critique of Google’s tangled position on the Web. *New Media & Society*, 16(2), 195–211. <https://doi.org/10.1177/1461444813481195>
- Ritzer, G., & Jurgenson, N. (2010). Production, consumption, prosumption: The nature of capitalism in the age of the digital “prosumer.” *Journal of Consumer Culture*, 10(1), 13–36. <https://doi.org/10.1177/1469540509354673>
- Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4), 990–1029. <https://doi.org/10.1162/154247603322493212>
- Rodriguez, C. (2001). *Fissures in the Mediascape. An International Study of Citizens’ Media*. Cresskill, NJ: Hampton Press
- Ruckenstein, M., & Schüll, N.D. (2017). The datafication of health. *Annual Review of Anthropology*, 46(1), 261–278. <https://doi.org/10.1146/annurev-anthro-102116-041244>

- Ruppert, E. (2018). *Sociotechnical Imaginaries of Different Data Futures. An Experiment in Citizen Data*. Erasmus University Rotterdam. <https://www.eur.nl/sites/corporate/files/2018-06/3e%20van%20doornlezing%20evelyn%20ruppert.pdf>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718820549>
- Sadowski, J., & Bendor, R. (2018). Selling smartness: Corporate narratives and the smart city as a sociotechnical imaginary. *Science, Technology, & Human Values*, 44(3), 540–563. <https://doi.org/10.1177/0162243918806061>
- Schermer, B.W., Custers, B., & van der Hof, S. (2014). The crisis of consent: How stronger legal protection may lead to weaker consent in data protection. *Ethics and Information Technology*, 16(2), 171–182. <https://doi.org/10.1007/s10676-014-9343-8>
- Schneider, I. (2018). Bringing the state back in: Big Data-based capitalism, disruption, and novel regulatory approaches in Europe. In: Sætnan, A.R., Schneider, I., & Green, N. (Eds.) *The Politics and Policies of Big Data: Big Data, Big Brother?* 129–175. Abingdon: Routledge.
- Schrock, A. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society*, 18(4), 581–599. <https://doi.org/10.1177/1461444816629469>
- Selwyn, N., & Pangrazio, L. (2018). Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data & Society*, 5(1), 1–12. <https://doi.org/10.1177/2053951718765021>
- Sharon, T. (2018). When digital health meets digital capitalism, how many common goods are at stake? *Big Data & Society*, 5(2), 1–12. <https://doi.org/10.1177/2053951718819032>
- Smith, E. (2015). Corporate imaginaries of biotechnology and global governance: Syngenta, Golden Rice, and corporate social responsibility. In: Jasanoff, S. & Kim, S-H. (Eds.) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, 254–276. Chicago, IL: University of Chicago Press.
- Snow, D.A., & Benford, R.D. (1988). Ideology, frame resonance and participant mobilization. *International Social Movement Research*, 1, 197–217.
- Solove, D. J. (2013). Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7), 1880–1903. <https://ssrn.com/abstract=2171018>
- Srnicek, N. (2017). *Platform Capitalism*. Cambridge, UK: Polity Press.
- Star, S.L., & Griesemer, J.R. (1989). Institutional ecology, ‘translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3), 387–420. <https://www.jstor.org/stable/285080>
- Statista (2019a). *Facebook – Statistics & facts*. <https://www.statista.com/topics/751/facebook/>
- Statista (2019b). *Google – Statistics & facts*. <https://www.statista.com/topics/1001/google/>
- Statista (2019c). *The 100 largest companies in the world by market value in 2019*. <https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/>
- Taylor, C. (2002). Modern social imaginaries. *Public Culture*, 14(1), 91–124. <https://doi.org/10.1215/08992363-14-1-91>
- Taylor, C. (2004). *Modern Social Imaginaries*. Durham: Duke University Press.
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), 1–14. <https://doi.org/10.1177/2053951717736335>
- Taylor, L. (2018). MyData’s Nordic model for data governance, a libertarian fantasy worth engaging with. *Global Data Justice Blog*, November 19. <https://globaldatajustice.org/2018-11-19-mydata-nordic-model/>
- 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. <https://doi.org/10.1177/0263775816663195>
- Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7). <http://firstmonday.org/article/view/4901/4097>

- Turow, J., Hennessy, M., & Draper, N. (2015a). *The Tradeoff Fallacy. How Marketers Are Misrepresenting American Consumers and Opening Them Up to Exploitation*. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2820060](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820060)
- Turow, J., McGuigan, L., & Maris, E.R. (2015b). Making data mining a natural part of life: Physical retailing, customer surveillance and the 21st century social imaginary. *European Journal of Cultural Studies*, 18(4–5), 464–478. <https://doi.org/10.1177/1367549415577390>
- Unger, R.M. (2007). *Free trade reimagined: The World Division of Labor and the Method of Economics*. Princeton, NJ: Princeton University Press.
- van Couvering, E. (2008). The history of the Internet search engine: Navigational media and the traffic commodity. In: Spink, A., & Zimmer, M. (Eds). *Web Search: Multidisciplinary Perspectives*, 177–206. Berlin: Springer.
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208. <https://doi.org/10.24908/ss.v12i2.4776>
- van Dijck, J., & Nieborg, D. (2009). Wikinomics and its discontents: A critical analysis of Web 2.0 business manifestos. *New Media & Society*, 11(5), 855–874. <https://doi.org/10.1177/1461444809105356>
- West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, 58(1), 20–41. <https://doi.org/10.1177/0007650317718185>
- Winner, L. (1978). *Autonomous Technology – Technics-out-of-control as a Theme in Political Thought*. Cambridge, MA: MIT Press.
- Winner, L. (1986). *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago, IL: University of Chicago Press.
- World Economic Forum (2011). *Personal Data: The Emergence of a New Asset Class*. <https://www.weforum.org/reports/personal-data-emergence-new-asset-class>
- Wyatt, S. (2004). Danger! Metaphors at work in economics, geophysics, and the Internet. *Science, Technology, & Human Values*, 29(2), 242–261. <https://doi.org/10.1177/0162243903261947>
- Yeung, K. (2016). ‘Hypernudge’: Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118–136. <https://doi.org/10.1080/1369118X.2016.1186713>
- Yousif, M. (2015). The rise of data capital. *IEEE Cloud Computing*, 2(2), 4. <https://doi.org/10.1109/MCC.2015.39>
- Zuboff, S. (1985). Automate / informate: The two faces of intelligent technology. *Organizational Dynamics*, 14(2), 5–18. [http://dx.doi.org/10.1016/0090-2616\(85\)90033-6](http://dx.doi.org/10.1016/0090-2616(85)90033-6)
- Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*. New York, NY: Basic Books.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30, 75–89. <https://doi.org/10.1057/jit.2015.5>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London: Profile Books.
- Zuiderent-Jerak, T. (2015). *Situated Intervention. Sociological Experiments in Health Care*. Cambridge, CA: MIT Press.
- Zwitter, A. (2014). Big data ethics. *Big Data & Society*, 1(2), 1–6. <https://doi.org/10.1177/2053951714559253>





## Original Publications

**Lehtiniemi, T. & Kortenesniemi, Y. (2017)**  
**Can the obstacles to privacy self-management be overcome?**  
**Exploring the consent intermediary approach**  
*Big Data & Society*, 4(2), 1–11



# Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach

Big Data & Society  
July–December 2017: 1–11  
© The Author(s) 2017  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2053951717721935  
journals.sagepub.com/home/bds  


Tuukka Lehtiniemi<sup>1</sup> and Yki Kortenesniemi<sup>2</sup>

## Abstract

In privacy self-management, people are expected to perform cost–benefit analysis on the use of their personal data, and only consent when their subjective benefits outweigh the costs. However, the ubiquitous collection of personal data and Big Data analytics present increasing challenges to successful privacy management. A number of services and research initiatives have proposed similar solutions to provide people with more control over their data by consolidating consent decisions under a single interface. We have named this the ‘*consent intermediary*’ approach.

In this paper, we first identify the eight obstacles to privacy self-management which make cost–benefit analysis conceptually and practically challenging. We then analyse to which extent consent intermediaries can help overcome the obstacles. We argue that simply bringing consent decisions under one interface offers limited help, but that the potential of this approach lies in leveraging the intermediary position to provide aides for privacy management. We find that with suitable tools, some of the more practical obstacles indeed can become solvable, while others remain fundamentally insuperable within the individuated privacy self-management model. Attention should also be paid to how the consent intermediaries may take advantage of the power vested in the intermediary positions between users and other services.

## Keywords

Personal data, informed consent, privacy self-management, privacy, cost–benefit analysis, consent intermediary

## Introduction

The production of personal data – any information relating to an identified or identifiable natural person (European Union, 2016) – seems to be ever increasing as activities performed with information technology have become daily routines, and companies use Big Data analytics to produce potentially detailed pictures of us. This extensive use of personal data can benefit individuals themselves, as personalisation can make services more valuable to use, and business models based on profiling often make services available free of charge. But the associated cost is the impact on privacy as people reveal more information about themselves to service providers.

EU legislation has long held privacy as a fundamental right of the individual (Wachter, 2017) and places strict limits on the processing of personal data. The new General Data Protection Regulation (GDPR) (European Union, 2016) states that personal data may

only be processed based on one of the following six grounds: it is required by a legal obligation, it is carried out to protect a vital interest of the individual, it is carried out for the public interest, it falls within a legitimate interest of the data controller, it is necessary for the performance of a contract, or it is based on the consent of the individual. The informed consent approach, which is also used in many other jurisdictions, allows people the freedom to agree to many types of data processing. However, with a contract, processing is limited to data which is strictly necessary for its fulfilment, and for the

<sup>1</sup>Department of Computer Science, Aalto University, Finland; Faculty of Social Sciences, University of Turku, Finland

<sup>2</sup>Department of Computer Science, Aalto University, Finland

## Corresponding author:

Tuukka Lehtiniemi, Aalto University, PO Box 15600, Espoo 00250, Finland.

Email: [tuukka.lehtiniemi@iki.fi](mailto:tuukka.lehtiniemi@iki.fi)



other bases, the individual can either do nothing or can at most object to some of it.

In this article, we focus exclusively on consent-based processing as it places the greatest demands on the individual's ability to make informed decisions. People are expected to manage their privacy by weighing the subjective costs and benefits of data collection in each case (Solove, 2013). In practice, however, many are neither well informed on the uses of their personal data nor feel in control of it (European Commission, 2015; Turow et al., 2015). A fundamental dilemma underlies the concept of informed consent: meaningful cost–benefit analysis on personal data is anything but straightforward in the context of Big Data analytics, data aggregation, and opaque data flows. But for the moment, we will live with the model of informed consent, as in many jurisdictions it is codified in legislation. This has sparked an ongoing discussion about how to make the model work better. For example, Custers (2016) discusses expiry dates for consents and calls for further discussion on the issue of consents in the Big Data era, a call to which we respond with the present article.

Within the last few years, a number of initiatives to give people better control over their personal data have started to appear. Proponents of personal information management systems (PIMS) (Abiteboul et al., 2015; European Commission, 2016) recognise the current inability of people to meaningfully control the uses of their data, and seek to redress the situation with personal data stores and features for managing data use permissions. We consider these emerging services *consent intermediaries* (CIs). With CIs, people themselves still manage their own privacy, but the intermediary consolidates all management to a single place. In this article, we investigate the concept of CIs and ask two questions: (1) to what extent can CIs help people in making informed privacy decisions and (2) is it even possible to overcome all of the obstacles?

The rest of the paper is organised as follows: we first review the privacy self-management model and identify eight obstacles which currently stand in the way of informed privacy decisions. We then proceed to describe the CI approach and analyse its potential to tackle these obstacles. We find that CIs form a platform for building tools to help with the more practical obstacles, but obstacles arising from the privacy self-management model's individuated nature are not as easy to solve without relaxing the model's individuated assumptions. Finally, we conclude by discussing the implications of the CI approach.

## Privacy self-management

The right to informational privacy is essentially a decision right. Zuboff (2015), for example, conceptualises it as the

ability to choose one's position along the spectrum between secrecy and transparency. Altman (1975) refers to the same phenomenon as boundary regulation of privacy and publicness. Solove (2013) refers to the current approach of privacy regulation as *privacy self-management*; people have the right to *notice* of the upcoming collection and use of personal data and have the *choice* whether or not to consent to such processing. Armed with these rights, people are expected to make privacy decisions based on cost–benefit evaluations and to disclose data only when the benefits outweigh the costs.

Privacy self-management has to take into account the highly divergent preferences people have on the desirable position along the secrecy–transparency spectrum. Westin's well-known classification identifies privacy *fundamentalists*, who have high privacy concerns, *pragmatists*, who have some concerns but favour individual choice, and the *unconcerned*, who have low concerns and tend to trust data collectors (Hoofnagle and Urban, 2014). Further, individuals' preferences on privacy can change over time, and are also highly context-dependent (Acquisti et al., 2015; Coll, 2014; Hoofnagle and Urban, 2014).

Privacy self-management relies on individuals being informed and making decisions based on subjective analysis of this information, and it therefore places a lot of faith in their rational capabilities. But in practice, decision-making is only partially the result of rational cost–benefit analysis. Decisions are also affected by social norms, emotions and heuristics (Acquisti et al., 2015), and people are only boundedly rational (Gigerenzer and Selten, 2001), due to limitations in information, cognitive capabilities, and available time. Therefore, individuals are often not that well-informed when consenting; they do not always read privacy policies (Custers, 2016) and can operate under misinformed assumptions about these policies' purpose and contents (Turow et al., 2015). In a recent Eurobarometer, only 18% of respondents reported reading privacy policies fully and 49% partially, length and complexity being typical reasons for not reading them (European Commission, 2015). In fact, many habitually accept consent dialogues without even glancing the provided information (Böhme and Köpsell, 2010).

Unsurprisingly, people do not feel in control of personal data but nevertheless see no alternatives to disclosing data in order to gain access to services (European Commission, 2015). The feelings of powerlessness to contest the data collection practices speak of the same issue (Andrejevic, 2014). To overcome the experienced lack of control, people employ *implicit* control mechanisms to regulate the quantity and quality of data, including maintaining multiple or pseudonymous profiles, providing incorrect information,

and refraining from providing data whenever feasible (Snell et al., 2012). Overall, however, people's actions seem to demonstrate the privacy paradox: despite indicating concerns about privacy, they part with intricate details about their private life – in other words, behavioural intentions towards privacy are not reflected in actual behaviour (Norberg et al., 2007). One explanation is that broad attitudes to privacy may measure different things than contextual decisions (Acquisti et al., 2015). Another could be that when people do consider the costs and benefits of the options provided, if the cost of achieving better privacy is high – for example the inability to use a social networking site, or a significant effort to configure the privacy settings – people do not always see the benefits of better privacy as worth the cost.

We can conclude that even if the privacy self-management model expects people to behave rationally, this is not always the case. It is therefore worth exploring whether the proposed new ways of consenting can help people make better privacy decisions.

## Obstacles to privacy self-management

In this section, we distil findings from literature review into eight obstacles which summarise the challenges of privacy self-management. Our aim is to categorise and reformulate these findings so that they can be used to evaluate the effectiveness of attempts to improve privacy self-management. The obstacles are summarised in Table 1.

**Timing and duration.** A challenge with privacy is that it is an outcome of long-term information management, but the practical implementations of privacy self-management do not currently support this (Solove, 2013). The point of decision occurs when the collection of personal data is started, and individuals are then expected to assess all future harms and benefits. Decisions on disclosing personal data are also made in isolation from other similar decisions, and often they are made with the aim of gaining immediate benefits. And while immediate harms may be insignificant, long-term harms can develop gradually over time. Having to make the decision before the outcomes arise is arguably a feature of most human decision-making, but with personal data, the timing poses particular difficulties due to the inherent dynamics arising from the advancement of data analysis technologies (Custers, 2016). As harms and benefits may arise by mechanisms which are not discernible, or do not yet even exist, the consequences of a disclosure decision are a moving target. Yet a consent, once given, is typically in effect indefinitely.

**Table 1.** Obstacles to privacy self-management.

Timing and duration	Estimating harms is difficult due to timing of decisions and the typically unlimited duration of the consent (Custers, 2016; Solove, 2013).
Non-negotiability	The terms are not negotiable enough (Custers, 2016).
Scale	Privacy self-management does not scale well enough (McDonald and Cranor, 2008; Solove, 2013).
Aggregation	Data is aggregated and analysed to produce new data, leading to implicit disclosure of latent data (Mai, 2016; Solove, 2013).
Downstream uses	Data flows to parties and purposes not foreseen at the time of consenting (Anthes, 2015; Crain, 2016; Turow et al., 2015).
Cognitive demands	The cognitive limitations of all human decision making hamper cost–benefit analysis (Solove, 2013).
Social norms	Pressure to conform can strongly affect the decisions people make (Acquisti et al., 2015; Andrejevic, 2014; Zuboff, 2015).
Social data	Privacy decisions are framed as individual choices, but the data and the decisions also affect others (Lampinen et al., 2011; Schneier, 2010; Taylor et al., 2017).

### Non-negotiability

The current implementation of notice and choice is usually based on terms dictated by the service provider (Custers, 2016), and users have to accept these terms in full to use the service. The other option, obviously, is not to use the service. This Hobson's choice does not match the preferences of those who are willing to agree to some subset of the terms in exchange for some subset of the service. Also, once the choice is made, the terms of personal data use are largely fixed. For example, the privacy settings within a service often only affect the visibility of personal data to third parties rather than, for example, what data gets collected. However, post-consent negotiations of sorts can arise when an organisation attempts to impose new terms which a large portion of the users find unacceptable. This is evidenced by the stir which Spotify's new privacy policy caused and the consequent changes made by the company (Kastrenakes, 2015). In addition, data protection regulation in the EU, for example, provides the possibility of withdrawing consent at will. But in practice, reconsidering a decision is impractical and potentially ineffective. When providing consent is an all-or-nothing decision, withdrawing consent involves ceasing the use of the service altogether. It also involves removing data

in the service provider's databases, but having data deleted has turned out to be a complex issue (Custers, 2016). Interestingly, the upcoming GDPR (European Union, 2016) addresses the current situation by stating that the availability of a service cannot be contingent on the individual consenting to data processing which is not essential to the service.

### Scale

A practical obstacle to decision-making is that privacy self-management as currently implemented does not scale too well. Making people better informed in their decisions cannot be achieved simply by convincing people to read privacy policies better (Solove, 2013), because there is just too much information to study, and there are too many decisions to make. An estimated 80–300 hours are needed to familiarise oneself with just the privacy policies of the websites an individual visits in a year (McDonald and Cranor, 2008), and including other data-collecting entities only increases the time required. Also, as we will discuss below, some of the obstacles are due to people not being fully aware of the complex consequences of the decisions they make – and the more people are made aware of the consequences, the more problematic the scaling problem can become.

### Aggregation

Data-collecting entities often aggregate personal data across individuals and contexts, which can lead to revelation of new data through data analysis. We contrast this *latent data* to openly expressed and exhaust data (Kitchin, 2014). Openly expressed data is consciously provided by individuals about themselves, for example filled in a form, and exhaust data is produced by observing activities, for example, clickstreams on a website. Latent data, however, is fundamentally different; it is produced from other data by using inference techniques and is therefore implicitly shared alongside other data. Yet an explicit consent is never provided for latent data (Mai, 2016). Inference can produce seemingly unconnected results by treating the input data as proxy data for the unavailable information. Then, for example, demographic data can be deduced from location history alone (Bellovin et al., 2013). Aggregation also happens across not just numerous sources of data but also across individuals. Therefore, the costs and benefits of my disclosure decisions are affected by the decisions of others. So even if each disclosure decision were well-considered in isolation, the aggregation of data can lead to the overall effect being undesired. The production of latent data also hinders the effectiveness of refraining from disclosing data, as it may still be deduced from other data (Custers, 2016).

### Downstream uses

Unexpected movements of personal data to new parties are to a large extent opaque to the individuals, complicating meaningful decision-making. We refer to these movements as downstream uses of data. There are multiple reasons for these movements. In downstream data markets, data brokers sell personal data compiled from public records and nonpublic sources, often without the knowledge of the individuals involved, even though they may have consented to such uses of data by the primary data collectors (Anthes, 2015; Crain, 2016; Turow et al., 2015). Changes in business models of data-collecting companies may result in new uses of the data contradicting the individual's expectations from the time of decision, an example being direct-to-consumer personal genome testing and the subsequent medical research use of the collected data (Alba, 2015; Seife, 2013). Another reason for downstream uses is malicious actions of third parties. Well-known examples include publication of data hacked (Zetter, 2013) or otherwise collected (Zimmer, 2016) from dating services, resulting in personal data about customers being put to unforeseen uses. Personal data collected by private companies can also end up being aggregated in governmental databases, and superficially innocuous pieces of personal data may end up being highly consequential in practice. The upcoming GDPR, again, places some limitations on downstream uses of data, stipulating that all processing must have a legal basis.

### Cognitive demands

People's ability to make informed and rational choices about personal data is not on par with requirements of privacy self-management, and people can end up making bad decisions with respect to disclosing personal data, regardless of the information and tools they have in use. The cognitive limitations hampering privacy self-management have been summarised by Solove (2013) as follows. To begin with, people are not very well informed about the decisions they make because they do not read privacy policies. If they do read them, they have difficulties understanding them. If they do understand them, they lack the necessary knowledge to make a truly informed choice. And even if they are well-informed, their decision-making capability is limited by difficulties which generally riddle human decision-making.

### Social norms

As observed above, the decision not to disclose personal data often means non-participation in activities which include collecting data. As many online services are regarded an integral part of modern life (Andrejevic, 2014; Zuboff, 2015), non-participation

may simply be infeasible regardless of privacy preferences or subjective concerns over data disclosure. Thus, decision-making on personal data is subject to social norms (Acquisti et al., 2015) which regulate individual decisions. Another way of saying this is that private cost–benefit decisions to disclose data are embedded in a network of social relations, and looking at them from an individuated, under-socialised point of view is misleading (Granovetter, 1985). These norms are further reinforced by each individual decision; the more people conform, the harder it becomes to deviate from the norm regardless of the individual judgement of costs and benefits. Norms may also be at odds with attempts to implicitly control the quality and quantity of data once consent has been provided.

### Social data

Privacy self-management frames the decision-making on personal data as an individual choice based on private cost–benefit analysis, despite personal data often also conveying information about others. Schneier (2010) uses the term *incidental data* to denote data which other people’s activities leak about you. Any data about my interactions or relationship with you is also data about you. Health or genome data may implicate relatives in the case of hereditary diseases, shared photos can convey information about others, and consumption data may by nature concern a household. Decision to share location data may help predict the future locations of others (Bellovin et al., 2013), and the combined effect of two people sharing location data may reveal details about their relationship. In particular, latent data produced by Big Data analytics may by nature concern a group rather than individuals (Taylor et al., 2017). Privacy can, by various mechanisms, be affected by the choices others make (Lampinen et al., 2011) and the outcomes of data-sharing decisions are, therefore, not only private.

To summarise the obstacles, privacy decisions are made in a situation described by considerable information asymmetry; non-experts know little about collected personal data, what is done with the data, or the business operations of the data industry (Zuboff, 2015). Altogether, the obstacles affect privacy self-management by first making it hard to appraise the situation and then by diminishing the possibilities of actually making preferred decisions.

Next, we proceed to describe the CI approach and analyse its potential to tackle these obstacles.

### Consent intermediaries

The last few years have seen an emergence of initiatives and services whose aim is to provide people with better

control over the collection and sharing of their personal data. A report by the European Commission (2016) on PIMS included commercial service developers such as the personal cloud server Cozy Cloud (2017) and the personal information control services digi.me (2017) and Meeco (2017), as well as research-originated initiatives such as the networked personal data indexing device Databox (Chaudhry et al., 2015), personal data stores Hub of All Things (Hub of All Things, 2017) and OpenPDS (de Montjoye et al., 2014), and the personal data management model MyData (Poikola et al., 2015).

All of these services aim to provide more control to personal data and allow people to share their data with third parties, but two different means to achieve this can be identified: *storage spaces* accumulate personal data from various sources, whereas *permission-management services* only keep track of where data is stored. While their practical implementations and stages of maturity vary, conceptually these services propose to act as Consent Intermediaries (CIs) between individuals and data-using entities. From the perspective of individuals, CIs aim to consolidate the provisioning of consents under one control point, providing an access point through which individuals grant, view and withdraw consent to collect and use data. From the perspective of the services, CIs enable the outsourcing of privacy management. The CI, therefore, consolidates the consenting practices of many services and the consenting decisions of multiple individuals, in a conceptual change to the current dispersed practice as shown in Figure 1.

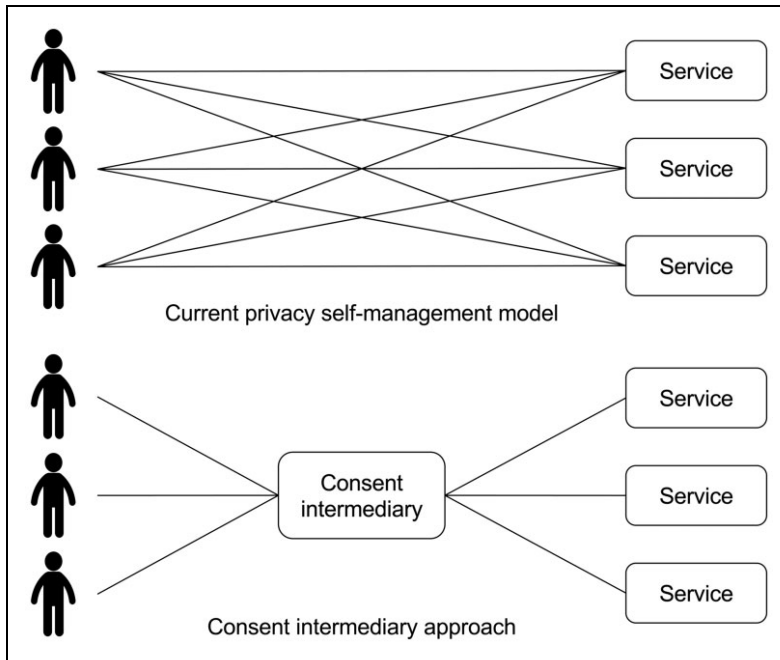
CIs strive to provide individuals with better control over their personal data, which is expected to lead to better privacy and larger benefits from their data. A recent opinion published by the European Data Protection Supervisor (2016), for example, sees the PIMS services, backed up by GDPR, as potentially leading to the empowerment of users. Yet empowerment and control might be illusory if making sense of the consequences of data disclosure decisions does not become easier than it currently is. In addition, CIs also introduce a new party in the consenting process, which may also have its own aims and incentives.

### Analysis

We start by looking at the changes the intermediary may bring to the consenting process, and then go through how these changes could help overcome the obstacles. We conclude the analysis by addressing the nature of privacy self-management obstacles.

#### *How could a consent intermediary change consenting?*

Simply bringing consents under one control interface has the potential to make privacy self-management



**Figure 1.** The conceptual change of the consent intermediary approach.

easier. It can help individuals make sense of the whole, be aware of past decisions, and take them into account in future decisions. This is particularly true if the CI presents consents in a comparable format. An overall view can help in situations in which privacy management fails due to individuals not being aware of the totality of their own decisions, which is, in light of the identified obstacles, a part of the problem but does not cover nearly all of its aspects.

Bigger changes can happen if the CI takes advantage of its intermediary position and builds new tools to aid decision-making, for example, by employing concepts which are already commonly used in other online services. Without an intermediary between individuals and data users, it would be much harder to build these tools.

First, online services and marketplaces routinely employ recommendations, predictions, ratings, and crowdsourcing to provide their users tailored information. In smartphone platforms, users give permissions to applications they install, and making privacy-conscious decisions requires accounting for how the applications likely use those permissions. Liu et al. (2016) propose a ‘privacy assistant’ which provides personalised recommendations for application permissions based on user profiling. It is possible to apply a similar approach to the more general issue of providing consent. An intermediary service can leverage consent metadata, including information contained in consents

themselves, and information on other users’ actions regarding consents to provide more information at the point of decision. In experimental settings, timely presentation of privacy information has been found to lead to more privacy-protecting decisions (Kelley et al., 2013), and designs which highlight the implications of decisions have been found to have a similar effect (Harbach et al., 2014).

Second, there are several ways to automate actions based on, for example, rules and profiles. With a CI, it might be possible to automate some practical consent decisions. This might include straightforward recommendations based on preferences; users indicate their preferences, and the intermediary then recommends actions based on them. Privacy preferences can also be deduced automatically using data analysis, as has been done for privacy settings on Facebook (Fang and LeFevre, 2010) and for mobile applications (Liu et al., 2016). In the context of mobile applications, recommendations for access permissions by experts (Rashidi et al., 2015) and crowdsourcing (Agarwal and Hall, 2013) have also been proposed. With consents, individuals could similarly choose to automatically follow the recommendations formed collaboratively by engaged users or provided, for example, by a privacy advocacy group or a commercial provider.

Third, companies are fundamentally dependent on individuals as their data sources, and this position



could be leveraged for more favourable terms of data use. Currently, the position of individuals is characterised by low bargaining power over these terms. Attempts to balance similar asymmetries are in many other contexts based on the disadvantaged parties organising as a collective actor rather than as individuals, including well-known examples of consumer interest lobbies and unions in labour market negotiations. The CI could act as a platform for collective action to balance these power asymmetries by leveraging the presence of others in the same decision-making situation.

### *How could these changes help overcome the obstacles?*

**Timing and duration.** As noted above, making information on consents viewable from a single point has the potential to increase individuals' awareness of the long-term aspects of privacy management. This could help individuals make decisions in a more systematic manner, particularly by mitigating the timing issue in the sense that the long-term effects of decisions can be better taken into account. Making sense of the whole can also be made possible by making use of consent metadata. For example, an individual might be made aware of all actors who have access to certain kind of data. It would be straightforward for the CI to employ nudges to revisit previous decisions (Liu et al., 2016) to see whether they still accurately represent current preferences. Prompts to re-evaluate consent might be issued periodically or be based on changed conditions such as the provision of new consent for similar purposes. Nudges and prompts could bring benefits similar to those of proposals for periodically expiring consents (Custers, 2016; Mayer-Schönberger, 2011), but would likely also exhibit similar problems; for some, they would likely be just another forced click of an 'agree' button without much thought (Custers, 2016).

**Non-negotiability.** Broad, non-negotiable consents make sense to many companies, as their business models drive them to make privacy policies as general as possible in terms of the quality, quantity and possible uses of personal data (Custers, 2016; Srnicek, 2017). Implementing negotiability of privacy policies, for example, by using smart contracts, may be costly in service design sense. Also, tweaking privacy policies before accepting them increases the decision-making effort required from individuals, and customised consents lead to the production of additional metadata and complicate data management (Custers, 2016). Despite the incentives for non-negotiable consent, there is nothing which fundamentally prohibits negotiations. To the extent that the lack of negotiations is attributable to each individual having low bargaining power against

data collectors, individuals could organise as a collective entity to leverage the dependence of organizations on them as data sources. We argue that such collective action is difficult to achieve without some kind of coordinating entity, and the CI could act as one. Introducing an intermediary between individuals and data-collecting entities would, in any case, affect the power balance of the situation, and in the best case this would help individuals have a say over the terms under which their data is used. However, it seems safe to assume that the intermediary might leverage its position also for its own benefit, which may or may not align with the interests of the individuals.

**Scale.** Given the amount of effort expected from individuals, the scale problem seems difficult to overcome. However, we argue that it is not a problem of principle but is largely due to how privacy self-management is implemented in practice. Making each decision simple enough, or gathering many small but similar decisions under one higher-level decision, would make the whole decision-making effort more manageable. Automation and aides which help identify important decisions would work in this manner. Practical questions include whether or not it is possible to simplify decisions enough while fulfilling both the expectations individuals have of the ability to affect each decision, and the requirements data protection regulations place on the way consent is provided.

**Aggregation.** Data aggregation is an obstacle which is difficult to tackle in principle, as latent data emerges only ex post, after decisions have been made. The relationship between disclosed data and the consequences of this disclosure are therefore obscured, and latent data can make meaningful analysis of costs and benefits impossible. While it is likely impossible to fully overcome this obstacle, some improvements to the current situation can be envisioned. Making people aware of known outcomes of data aggregation, based not only on their own but also on others' past decisions, could help them become better informed about potential latent data. Consent metadata can, for example, be aggregated across individuals and used to provide information on what data others have chosen to disclose. Simply explaining the likely purpose of a mobile application's permission request has been found to play an important role in privacy decisions (Liu et al., 2016). Metadata can also be used to form predictions of likely consequences of disclosure decisions without access to the actual disclosed data, for example, predicting the potential revealing of contextual data based on location data alone, if others have already provided contextual and location data (Bellovin et al., 2013). Such consequences are, of course, a moving target, and predictions

would necessarily be coarse, but they might still be better than the heuristics individuals currently have to rely on.

**Downstream uses.** To the extent that downstream uses of data happen with user's consent, it is possible in principle to make it easier to take these uses into account in privacy decisions. For example, by combining consent metadata with other data sources, the network of data flows originating from the initial data collectors could be tracked in a manner similar to tracing the relationships of online ad platforms (Helmond and van der Vlist 2016), and visualising them might make sense-making easier. Naturally, the possibility of increasing transparency is limited to data flows which are consented to and trackable – excluding, for example, downstream uses through surveillance or data leaks. In addition, structural constraints of the data industry, such as opaque business practices and analytical layers which separate data sources from data uses, limit efforts to increase transparency (Crain, 2016).

**Cognitive demands.** The cognitive limits of human decision-making fundamentally restrict cost–benefit analyses. An obvious way to tackle this problem is to make decisions less demanding. On-time provision of relevant information could make it less demanding to be informed, but all efforts to make now-opaque consequences of data disclosure transparent run the risk of making each decision even more complex. Here, as in the context of the scale problem, one solution facilitated by a CI is to change the nature of the decisions; instead of considering each decision separately, an overall decision could be made on privacy management principles. The CI would offer a limited choice of more or less conservative privacy profiles, and then would recommend actions based on those profiles. At the extreme end, technically nothing prevents totally automated consenting, so that the intermediary would automatically provide consent on behalf of the individual or revoke consent from services no longer in use. Such solutions, however, may be at odds with current privacy regulations.

**Social norms.** The adherence to social norms makes an individual's privacy decisions dependent not only on their private costs and benefits, but also on others' expectations about those decisions. Tools which help evaluate the consequences of disclosure affect the private aspects of decisions to disclose data, and to the extent that social norms regulate those decisions, tools do not help. The obstacle that norms place in the way of privacy self-management is, therefore, insuperable within the individuated model, regardless of the privacy management tools developed. Of course, it

should be kept in mind that norms with respect to the disclosure of data are not fixed, and they may change over time.

**Social data.** There is a fundamental inconsistency between privacy self-management and the social nature of personal data. Social data makes my privacy dependent on the choices of others (and vice versa). My goals and privacy preferences might be contradictory to those of others, and the private benefits someone draws from disclosing social data might, from their perspective, overcome the private costs imposed on others. While the interdependencies of decisions and the consequences of my decisions on others can be made more visible by using tools similar to those discussed above, no amount of awareness will solve this fundamental tension.

### *How difficult are the obstacles?*

A key dilemma in all the discussed improvements to privacy self-management is that tools should help individuals take more complexity into account, and at the same time render decision-making easier. By revealing more of the consequences of data processing, we make the individual better informed, but this also makes decisions cognitively more demanding. Therefore, it seems to us that progress could best be made if CIs provided privacy management features on all of the three fronts described above. While not all the privacy self-management obstacles can be overcome, evaluation aides, decision automation and collective action have the potential to lead to better privacy self-management.

Based on our analysis, we can also deduce something about the nature of the privacy self-management obstacles. Some of them seem to be more practical in nature, and potentially solvable by developing tools for privacy management. Timing and duration, non-negotiability and the scale problem can, in principle, be solved by rethinking practices and providing new kinds of privacy management tools. While we consider these problems to be solvable in the sense that they are practical, it does not mean that they are easy to solve. At the other end of the spectrum, obstacles which feature social dimensions exhibit fundamental tensions with the individuated privacy self-management model and are therefore insuperable, unless the individuated principle of the model itself is changed. In between are the cognitive demands of decision-making, aggregation, and downstream uses of data. Privacy management tools can help to mitigate them, but they are challenging issues and exhibit aspects which we consider likely to be unsolvable. Table 2 presents this rough, by necessity, categorisation.

**Table 2.** Potential to overcome obstacles with privacy self-management tools.

<b>Solvable</b>	
Timing and duration	Practical problem of making it feasible to revisit decisions and revoke consent.
Non-negotiability	Practical problem of negotiating power.
Scale	Practical problem of making each decision easy enough.
<b>Challenging</b>	
Aggregation	Possible to mitigate by increasing awareness of latent data.
Downstream uses	Possible to mitigate by providing information on consented and traceable data flows.
Cognitive demands	Possible to mitigate by changing the nature of decisions.
<b>Insuperable</b>	
Social norms	Cannot be overcome within the individuated model.
Social data	Cannot be overcome within the individuated model.

## Discussion

Our overview of the potential of the CI approach was largely positive in nature, in that we looked at the possibilities of developing features which are potentially beneficial for individuals. It is clear that the restrictions of the approach, the implications of the CIs, and the limitations of privacy self-management also merit discussion.

To begin with, the CI approach rest on the assumption that people are inclined to manage privacy. While the experienced lack of control and implicit means used to gain some control exhibit a demand for better privacy self-management tools, some might simply be happy with current services. It is likely that new tools for privacy management alone will not overcome disinterest.

The existence of economic incentives to maintain the current state of affairs should not be overlooked. The production of latent data is ingrained in the business models of many online companies (Srnicek, 2017; Zuboff, 2015); therefore, it is one underlying reason for the extensive collection of personal data in the first place. If the privileged position of organisations which collect and use data is an outcome of privacy self-management (Coll, 2014), then the current scattering of consents and the associated difficulties in privacy self-management serve existing business interests. Attempts to change the existing consent practices are, therefore, likely met with resistance. Here, legal developments such as GDPR can have a significant impact.

For several of the privacy management obstacles, it is evident that the problem is connected to the

fundamental assumption of the privacy self-management model: that individuals themselves consider costs and benefits of data disclosure case by case. As outlined above, these problems could be managed by automating or delegating decisions. The level of abstraction can be increased, and the decision would then concern the rules of automation or to whom the decision would be delegated. Therefore, automation and delegation of consenting decisions could well lead to a better outcome, on the whole. The extent to which these actions are possible within current regulatory contexts is an object of research in its own right.

This assumption also makes privacy self-management an inherently individuated model. The social dimensions of personal data are hidden by framing the issue as ‘my data’ which is ‘about me’ (Crabtree and Mortier, 2015). This framing leaves open the question of social data which is not only about me, and focusing on individual cost–benefit analysis downplays the role of the norms which affect decisions. While it can be possible to make the social and societal consequences of data disclosure decisions more transparent, private cost–benefit analyses can still fail to take common good into account, which has the risk of leading to only locally optimal solutions; sometimes taking a broader societal or collective view may lead to a better, globally optimal solution. It would therefore be misleading to think about the social obstacles to privacy management as ‘problems’ which new consent practices can ‘solve’. Rather, they are features of the privacy self-management model and present an inherent tension which cannot be overcome without changing the underlying individuated principle of the model. The obvious question, then, becomes how to take the collective aspect into account in the relationships between individuals and data collectors. Concepts such as networked privacy (Lampinen, 2015) and focusing on the group rather than the individual as the starting point of privacy (Taylor et al., 2017) pave the way towards such alternative models. While this conceptual discussion is ongoing, practical experiments in collective privacy management are underway in more limited contexts, for example, developing an extended notion of ownership of digital content and providing tools which help in reaching collective decisions regarding such content (Squicciarini et al., 2009). It might be possible to extend solutions like this to the more general context of consent as well, which would amount to developing a model to achieve common good from an individual privacy management starting point. CIs can likely function as platforms which facilitate the building of tools from these wider points of view as well. However, the individuated approach required by privacy regulations might render such solutions non-compliant.

While forms of automation could lead to easing the cognitive load of decision-making, automation does not come without its own trade-off: decision-making power is transferred to those forming profiles and recommendations, such as algorithm designers. More generally, given the power invested in intermediary positions between user and data-using entities, we should pay close attention to how intermediaries make use of this power. Intermediaries could channel collective action and set up governance mechanisms which participating organisations are expected to follow, which could well work favourably for individuals. But we should not assume this is the only possible outcome. The intermediary also has the capacity to affect the behaviour of its users, for example through discreet nudges or outright limits to choices. This leads to the possibility of coaxing users towards behaviours which serve its own ends. This also renders the intermediaries tempting targets for attacks, both for the troves of information they contain about the individuals in the form of consents, profiles, and policies, and for the power of influencing the individuals in their privacy decisions.

## Conclusions

As we will live with the consent-based privacy self-management model for some time, it pays to investigate ways to make it better. From the recent developments of personal data services, we identified the concept of CIs which gather privacy decisions under a single control point. Based on our analysis, this provides only some direct remedies to the obstacles which currently hinder privacy self-management. However, intermediaries could be leveraged to develop tools to mitigate obstacles, helping people understand the decisions they make, better evaluate their consequences, and simplify the decisions themselves. We conclude that it is indeed possible to make privacy self-management work better, and some of its obstacles seem to be even solvable with new tools. However, not all of the obstacles can be tackled this way. Some obstacles seem challenging in the sense that they could be only mitigated but likely not solved. Finally the inherent problems related to individual-centricity of the model lead to insuperable problems that could be better approached if its individuated assumptions were relaxed.

## Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their constructive feedback on this article.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

- Abiteboul S, André B and Kaplan D (2015) Managing your digital life. *Communications of the ACM* 58(5): 32–35.
- Acquisti A, Brandimarte L and Loewenstein G (2015) Privacy and human behavior in the age of information. *Science* 347(6221): 509–514.
- Agarwal Y and Hall M (2013) ProtectMyPrivacy: Detecting and mitigating privacy leaks on iOS devices using crowdsourcing. In: *Proceeding of the 11th annual international conference on mobile systems, applications, and services*, Taipei, Taiwan, 25–28 June, pp.97–110. New York: ACM.
- Alba D (2015) 23andMe teams with Big Pharma to find treatments hidden in our DNA. *Wired*.
- Altman I (1975) *The Environment and Social Behavior. Privacy–Personal Space–Territory–Crowding*. Monterey: Brooks-Cole Publishing Company.
- Andrejevic M (2014) The Big Data divide. *International Journal of Communication* 8: 1673–1689.
- Anthes G (2015) Data brokers are watching you. *Communications of the ACM* 58(1): 28–30.
- Bellovin SM, Hutchins RM, Jebara T, et al. (2013) When enough is enough: Location tracking, mosaic theory, and machine learning. *NYU Journal of Law & Liberty* 8: 555–628.
- Böhme R and Köpsell S (2010) Trained to accept? A field experiment on consent dialogs. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Atlanta, USA, 10–15 April, pp.2403–2406. New York: ACM.
- Chaudhry A, Crowcroft J, Howard H, et al. (2015) Personal data: Thinking inside the box. *Aarhus Series on Human Centered Computing* 1(1): 29–32.
- Coll S (2014) Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. *Information, Communication & Society* 17(10): 1250–1263.
- Cozy Cloud (2017) Cozy cloud website. Available at: <https://cozy.io/en/> (accessed 20 April 2017).
- Crabtree A and Mortier R (2015) Human data interaction: Historical lessons from social studies and CSCW. In: *ECSCW 2015: Proceedings of the 14th European conference on computer supported cooperative work*, Oslo, Norway, 19–23 September, pp.3–21. Cham: Springer.
- Crain M (2016) The limits of transparency: Data brokers and commodification. *New Media & Society*. Epub ahead of print 7 July 2016. DOI:10.1177/1461444816657096.
- Custers B (2016) Click here to consent forever: Expiry dates for informed consent. *Big Data & Society* 3(1): 1–6.
- de Montjoye YA, Shmueli E, Wang SS, et al. (2014) openPDS: Protecting the privacy of metadata through SafeAnswers. *PLoS one* 9(7): 1–9.
- Digi.me (2017). Digi.me website. Available at: <https://digi.me> (accessed 20 April 2017).
- European Commission (2015) Data protection. Special Eurobarometer (431).

- European Commission (2016) An emerging offer of 'Personal Information Management Services' – Current state of service offers and challenges. European Commission Report.
- European Data Protection Supervisor (2016) EDPS opinion on personal information management systems. Towards more user empowerment in managing and processing personal data. Opinion 9/2016.
- European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council. *Official Journal of the European Union* L119: 1–88.
- Fang L and LeFevre K (2010) Privacy wizards for social networking sites. In: *Proceedings of the 19th international conference on world wide web*, Raleigh, USA, 26–30 April, pp. 351–360. New York: ACM.
- Gigerenzer G and Selten R (2001) Rethinking Rationality. In: Gigerenzer G and Selten R (eds) *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press, pp. 1–12.
- Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91(3): 481–510.
- Harbach M, Hettig M, Weber S, et al. (2014) Using personal examples to improve risk communication for security & privacy decisions. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Toronto, Canada, 26 April–1 May, pp.2647–2656. New York: ACM.
- Helmond A and van der Vlist FN (2016) Big Data advertising infrastructures: A comparative study of social media Ad platforms. In: *Internet, Politics & Policy 2016*. 22–23 September, Oxford, UK.
- Hoofnagle CJ and Urban JM (2014) Alan Westin's privacy homo economicus. *Wake Forest Law Review* 49(261): 261–317.
- Hub of All Things (2017) Hub of All Things GitHub page. Available at: <https://github.com/Hub-of-all-Things> (accessed 20 April 2017).
- Kastrenakes J (2015) Spotify updates privacy policy with clearer language after backlash. *The Verge*.
- Kelley PG, Cranor LF and Sadeh N (2013) Privacy as part of the app decision-making process. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Paris, France, 27 April–2 May, pp.3393–3402. New York: ACM.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.
- Lampinen A (2015) Networked privacy beyond the individual: Four perspectives to "Sharing". *Aarhus Series on Human Centered Computing* 1(1): 1–4.
- Lampinen A, Lehtinen V, Lehmuskallio A, et al. (2011) We're in it together: Interpersonal management of disclosure in social network services. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Vancouver, Canada, 7–12 May, pp.3217–3226. New York: ACM.
- Liu B, Andersen MS, Schaub F, et al. (2016) Follow my recommendations: A personalized privacy assistant for mobile App permissions. In: *Proceedings of the 12th symposium on usable privacy and security*, Denver, USA, 22–24 June, pp.27–41. Berkeley: USENIX.
- McDonald AM and Cranor LF (2008) The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4(543): 1–22.
- Mai JE (2016) Big data privacy: The datafication of personal information. *Information Society* 32(3): 192–199.
- Mayer-Schönberger V (2011) *Delete: The Virtue of Forgetting in the Digital Age*. Princeton: Princeton University Press.
- Meeco (2017) Meeco website. Available at <https://meeco.me> (accessed 20 April 2017).
- Norberg PA, Horne DR and Horne DA (2007) The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs* 41(1): 100–127.
- Poikola A, Kuikkaniemi K and Honko H (2015) *MyData – A Nordic Model for Human-Centered Personal Data Management and Processing*. Helsinki: Finnish Ministry of Transport and Communications.
- Rashidi B, Fung C and Vu T (2015) Dude, ask the experts! Android resource access permission recommendation with RecDroid. In: *2015 IFIP/IEEE international symposium on integrated network management (IM)*, Ottawa, Canada, 11–15 May, pp.296–304. New York: IEEE.
- Schneier B (2010) A taxonomy of social networking data. *IEEE Security and Privacy* 8(4): 88.
- Seife C (2013) 23andMe is terrifying, but not for the reasons the FDA thinks. *Scientific American*.
- Snell K, Starkbaum J, Lauß G, et al. (2012) From protection of privacy to control of data streams: A focus group study on biobanks in the information society. *Public Health Genomics* 15(5): 293–302.
- Solove DJ (2013) Privacy self-management and the consent dilemma. *Harvard Law Review* 126(7): 1880–1903.
- Squicciarini AC, Shehab M and Paci F (2009) Collective privacy management in social networks. In: *Proceedings of the 18th international conference on world wide web*, Madrid, Spain, 20–24 April, pp.521–530. New York: ACM.
- Srnicek N (2017) *Platform Capitalism*. Cambridge, UK: Polity Press.
- Taylor L, Floridi L and van der Sloot B (2017) Introduction: A new perspective on privacy. In: Taylor L, Floridi L and van der Sloot B (eds) *Group Privacy. New Challenges of Data Technologies*. Cham: Springer International Publishing, pp. 2–12.
- Turow J, Hennessy M and Draper N (2015) The Tradeoff Fallacy. How marketers are misrepresenting American consumers and opening them up to exploitation. A report from the Annenberg School for Communication. University of Pennsylvania.
- Wachter S (2017) Privacy: Primus inter pares – Privacy as a precondition for self-development, personal fulfilment and the free enjoyment of fundamental human rights. Available at: <https://ssrn.com/abstract=2903514> (accessed 20 April 2017).
- Zetter K (2013) Hackers finally post stolen Ashley Madison data. *Wired Magazine*.
- Zimmer M (2016) OkCupid study reveals the perils of Big-Data science. *Wired Magazine*.
- Zuboff S (2015) Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30: 75–89.



**Lehtiniemi, T. (2017)**  
**Personal data spaces: An intervention in surveillance capitalism?**  
*Surveillance & Society*, 15(5), 626–639





Tuukka Lehtiniemi

Department of Computer Science, Aalto University, Finland  
Department of Social Research, University of Turku, Finland  
tuukka.lehtiniemi@iki.fi

---

## Abstract

Personal data spaces, or PDSs, are emerging intermediary services that allow users control over the sharing and use of their data. In this article, the surveillance capitalism model, which describes how businesses employ datafication to create value in the digital economy, is used to contextualize PDSs. Focusing on three PDS services, I analyze the social imaginaries they represent, paying attention to the increased agency over data they offer users. This proposed agency reflects the efforts of PDSs to intervene in, but not counter, surveillance capitalism. While their goal is to intensify datafication by increasing the quality and specificity of data that businesses can employ, their interventions also change the structure of data flows, allowing users to more directly benefit from datafication. PDSs envision their users as data-supplying and benefit-demanding market participants, active subjects in value creation instead of passive objects of data extraction. PDSs view themselves as platform providers that facilitate data exchanges and rely on market mechanisms to ensure beneficial services are developed for users to choose from.

---

## 1. Introduction

In the digital economy, value creation relies on *datafication* (Mayer-Schönberger and Cukier 2013): the transformation of aspects of people's lives into quantified data. A stream of research has explored the connections between datafication, surveillance, and monetization of data. For example, van Dijck (2014) discusses how businesses employ data to monitor and monetize behavior online. Exploring power asymmetries related to datafication, Andrejevic (2014) points out the differences in capability between those who collect and mine data and those whom the collection targets. Following similar arguments, Andrejevic and Gates (2014) note the systemically opaque nature of data analytic processes, and Crain (2016) stresses the significance of unilateral control of the conditions of commodification of data.

Zuboff (2015) discusses the evolution of computer mediation from the workplace into the online space and argues that datafication has given rise not only to new opportunities to learn about those whom data concerns, but also to new contests over learning. The pertinent questions have become who can learn, how, and what—and particularly, who decides about these things. Zuboff argues that to understand how the answers to these questions about learning are shaped, the underlying model of value accumulation must be understood. Zuboff calls the institutionalized value accumulation model in the online space *surveillance capitalism*: a specific form of informational capitalism (Castells 1996) pioneered by Google and employed by large online companies as well as online startups. Surveillance capitalism monetizes data

Lehtiniemi, Tuukka. 2017. Personal Data Spaces: An Intervention in Surveillance Capitalism? *Surveillance & Society* 15(5): 626-639.

<http://library.queensu.ca/ojs/index.php/surveillance-and-society/index> | ISSN: 1477-7487

© The author(s), 2017 | Licensed to the Surveillance Studies Network under a [Creative Commons Attribution Non-Commercial No Derivatives license](#).

acquired through surveillance. It operates on data extracted from users, turns extracted data into behavioral predictions, and often monetizes them through markets that users cannot participate in. The taken-for-granted assumptions about how this is done already shape the answers to questions about learning.

Discontent with the interrelationship between datafication and value creation has prompted new initiatives from technology developers, who promise to empower people to take control of processing of personal data (European Data Protection Supervisor 2016; Poikola et al. 2015). These initiatives include emerging intermediary services (Abiteboul et al. 2015; European Commission 2016) that provide users with a personal data space, or PDS, referring to a data storage service coupled with interfaces to manage flows of data. PDSs promise users a place in the driver's seat for the uses of their personal data (Spiekermann et al. 2015): users would make decisions on how, with whom, and for what purposes their data are shared. I consider PDSs to be representations of social imaginaries (Kelty 2008; Taylor 2004) of how the data economy should work. As I will discuss in this article, in these imaginaries people are able to reap more of the benefits of datafication for themselves by directing data only to uses they deem individually beneficial. Hence, PDSs appear to promote a new capacity to act towards data.

Attempts to turn people from data sources into active data subjects resonate with the critical scholarship on datafication. In particular, they seem to tie in with Zuboff's contestation over learning. At the superficial level at least, PDSs seem to propose an intervention in the current ways companies make use of datafication to learn about users and to predict and modify their behavior. The purpose of this article is to explore this intervention in the value accumulation model and the unilateral market operations of surveillance capitalism. Specifically, I approach these issues with two interconnected research questions. First, what agency towards data do PDSs offer people? Second, how, as a consequence, do they propose to transform the economic role of people in value creation from personal data?

In Section 2, I look more closely at the value creation model of surveillance capitalism. Section 3 situates PDSs with initiatives for empowerment in the context of datafication. In Sections 4 and 5 I describe three PDSs, focusing on the agency towards data they propose. Section 6 discusses their intended intervention in surveillance capitalism, and Section 7 concludes with observations on issues that remain open.

## 2. Datafication and Surveillance Capitalism

Datafication as a basis for value creation is perhaps nowhere as obvious as in the context of online platforms. Platform companies like Facebook and Google offer services for free to consumers and expect profits from customers in other markets, often including markets where they sell targeting to advertisers. Gillespie (2010) describes how the strict computational understanding of platforms as infrastructure enabling the deployment of applications has been relaxed to favor a more everyday understanding of platforms as online services of various intermediaries. From a theoretical stance, Gawer (2014) identifies two perspectives to platforms: from an engineering perspective, platforms are modular technological architectures, and from an economic perspective, they are intermediaries of multi-sided markets. From the economic perspective, a company operating the platform creates products or services that facilitate exchanges between different types of market participants (Evans and Schmalensee 2011). Provision of free services to a group of customers is not an exceptional feature of online platforms. Platform companies often subsidize losses incurred in some sides of the market in order to stimulate sales in other, profit-turning sides (Rochet and Tirole 2003).

Google search provides a helpful example of how the platform logic works online. Rieder and Sire (2013) identify three distinct parties whose interactions the search platform mediates: users, content providers, and advertisers. Interactions between these parties take place in two markets. In the consumer market, the search service allows users and content providers to meet. In the other market, Google sells targeting to

advertisers. Targeting is based on data collected about the users as they use Google's services, and advertisements are displayed to users beside search results. Google has incentives to influence users' actions in the consumer market in ways that help maximize revenue from advertisers in the other market (Rieder and Sire 2013). Similarly, social media platforms have also been analyzed as multi-sided market intermediaries (Helmond 2015). In this case, the platform makes users, content providers, advertisers, and application developers meet.

Similarities between value creation models online are no coincidence. Zuboff (2015) argues that the market economy in general tends to gravitate towards dominant models for value creation, which eventually become the institutionalized, taken-for-granted context in which companies operate. In the online space, Zuboff argues, *surveillance capitalism* has emerged as the dominant model. In surveillance capitalism, companies aim to produce 'objective and subjective data about individuals and their habitats for the purpose of knowing, controlling, and modifying behavior to produce new varieties of commodification, monetization, and control' (2015: 85). The multi-sided markets operated by online platforms are practical instances of this model, and its assumptions are embedded into the ways in which these companies collect, store, and use data about their users.

### *Value Creation In Surveillance Capitalism*

Surveillance capitalism encompasses mass dataveillance—the systematic monitoring of people, by means of personal data systems, in order to regulate or govern their behavior (Clarke 1988; Degli Esposti 2014)—within a value creation logic. Value creation in surveillance capitalism is based on extracting data about users, analyzing these data to produce behavioral predictions, and monetizing these predictions by means of prediction products such as targeting and personalization (Zuboff 2015, 2016). Zuboff views this model as a continuation of developments she observed in increasingly computer-mediated work environments starting from the 1980s: the capacity of information technology to produce information on what it automates (Zuboff 1985) gave rise to new contentions and divisions concerning the ability to learn things based on information, and the power to decide who gets to do this learning. Today, similar power dynamics are present not only in the workplace, but generally in the online space, where the answers to the questions of who gets to learn and who decides about learning are shaped by the underlying value creation model.

Obviously, Zuboff is not alone in observing similarities in models of value creation from data. For example, van Dijck (2014) describes the 'big data mindset' of social media platforms in terms of measuring, manipulating, and monetizing human behavior, and Srnicek (2017) details the overall role of platforms in the contemporary economy. Zuboff's (2015) description of surveillance capitalism, however, is nuanced in the users' role, which makes it fitting for the purposes of analyzing personal data spaces. Towards this end, I highlight features of data extraction, decisions on data, production of predictions, and monetization of data that are pertinent to users' participation in these processes.

*Data extraction.* Zuboff asserts that data extraction in surveillance capitalism is a one-way process that occurs in the absence of dialogue between companies and their users, despite data signaling personal and potentially intimate details about users. This lack of reciprocity is supported by a number of observations on the demands for scale and scope of data extractive processes. Due to the probabilistic nature of their analytic capabilities, surveillance operations primarily value the quantity of data (Zuboff 2015). The extraction of data is not selective; all possible data on users' actions are considered signals to be analyzed (Mayer-Schönberger and Cukier 2013), and as much data as possible is recorded in order to determine their usefulness later (Andrejevic and Gates 2014). Accordingly, the extraction of data increasingly takes place beyond the immediate boundaries of online platforms, decentralizing their capacity to extract data about their users to the open internet (Gerlitz and Helmond 2013; Helmond 2015). These tendencies are explained by network effects: if service quality can be improved by data analysis, extracting more data leads to more users choosing the service, which again leads to better service (Rieder and Sire 2013;

Srnicek 2017). This incentivizes companies to broaden the scale and scope of data extraction. The role of users, then, is to be the source of as much data as possible.

*Decision rights.* The legal approach to protecting informational privacy is based on providing people with rights to be notified about data collection and to make a choice about it (Solove 2013). This means data extraction requires asking for user consent. That companies can extract data would, then, indicate that users have grown accustomed to trading data in exchange for services (van Dijck 2014). However, asking for user consent is contextualized by an asymmetrical relationship, in which the terms of data extraction are imposed on the users (Degli Esposti 2014). Privacy scholars (e.g., Acquisti et al. 2015; Solove 2013) highlight various shortcomings in the way privacy rights are enacted: users provide consent for data extraction in conditions characterized by lack of transparency, context-dependent and malleable attitudes towards privacy, and un- or misinformed decisions regarding disclosure of data. In the context of ubiquitous data extraction and advanced data analytics, then, practical possibilities for users to provide meaningful consent are limited. Zuboff (2015) asserts that by asking users to provide broad consent for extracting and using data, companies have in fact been able to gain decision rights over data for themselves.

*Production of predictions.* By using analytics on data extracted from users, companies produce behavioral predictions about the users: for example, their intentions, characteristics, or preferences. Producing predictions requires specialized means of production that rely on proprietary knowledge and capabilities (Zuboff 2015). Even if users are aware of predictions, they have only limited opportunities to view or correct them, and limited access to the information needed to comprehend the process producing them. The control of means of behavioral prediction, then, is asymmetric. This asymmetry, described as the ‘big data divide’ by Andrejevic (2014), further institutionalizes the lack of reciprocities between the company and users. It also gives companies the possibility to exercise ‘calculative power’ (Callon and Muniesa 2005): companies can assess the value of data extracted from users, and simultaneously limit the users’ possibilities of performing the same valuation, which also limits their economic action towards data. Further, predictions are also employed to modify behavior, for example by constructing personalized choice environments that do not necessarily enforce or restrict choices, but rather nudge users towards preferred outcomes (Yeung 2016). The production of predictions, then, is characterized by asymmetries arising from differences in access to the capabilities of data collection and analysis.

*Monetization.* In the end, who can make use of behavioral prediction and modification is determined in the market at which predictions are monetized. These markets are largely constructed by the companies. Rieder and Sire’s (2013) micro-level analysis of Google’s incentives to organize its markets in a self-serving way shows an example of this in practice. The markets for prediction products most famously face advertisers—importantly, users do not generally participate in transactions in these markets (Zuboff 2015).

To summarize, in Zuboff’s surveillance capitalism, the role of users in different stages of the process of value creation is largely characterized by a lack of reciprocities. Companies are able to exert significant control on data extraction and the production of predictions, and are able to shape the conditions of monetization of data. Surveillance capitalism’s response to the question of who decides what data are collected, what is learned based on it, and who does the learning, is decidedly ‘not people’.

### 3. Initiatives Aiming for User Empowerment

#### *Empowering Consumers and Citizens*

Discontent over the role users play in surveillance capitalism has given rise to various initiatives that offer people new capacities over data. Demands for transparency of the uses of data (Crain 2016; Richards and King 2014) are one example. Crain (2016) observes that transparency is a prevailing theme of consumer

empowerment online. He examines the data broker industry and argues that transparency runs into structural constraints arising from the political economy of commercial surveillance. Consumers are separated from the actual buyers and sellers of data by complex market arrangements that defy meaningful transparency. Moreover, much of the data the industry handles is separated from consumers by an analytical layer: while consumers are the source of raw data, the computationally generated predictions do not have a direct empirical source in consumers (Mai 2016). Crain (2016) considers the initial commodification of personal data to be at the root of power imbalances, and concludes that projects for consumer empowerment are toothless as long as commodification of data is taken for granted. According to Crain, empowerment through transparency is, then, bound to be unsuccessful.

The open data movement (Baack 2015; Chignard 2013; Kitchin 2014) presents another example, focusing on citizen empowerment. Baack (2015) investigates the movement as a reaction to the uneven distribution of power and knowledge due to datafication. He describes how open data activists consider the distribution to favor companies and governments, and how this, in turn, hinders public agency. Open data activists regard the availability of raw data as a prerequisite for generating knowledge. Therefore, in their view, the interpretive monopolies of raw data holders could be broken by making data openly available. By means of utilizing open data, say the activists, everyone could make their own interpretations, instead of relying on the interpretations of others. On one hand, then, the activists criticize how datafication leads to monopolization of interpretation. On the other hand, their goal is to turn datafication to support, instead of hinder, public agency. The activists, according to Baack, acknowledge that in order to make new interpretations possible, simply opening up the sources of raw data is not enough; empowering intermediaries that act between people and data-holding institutions are needed also.

### *Personal Data Spaces*

Personal data spaces, or PDSs, are another reaction to observed issues with the users' role in surveillance capitalism. They are intermediary services allowing users to store personal data and control their sharing with third parties. As I will discuss below, they have similarities with transparency and open data initiatives in their objective to provide people with new capacities with respect to data. PDSs have spurred both commercial and policy interest exemplified by a recent report of the European Commission (2016) which included over 20 'personal information management systems' from private-sector developers, academia, and nonprofits, and by the MyData 2016 conference (MyData 2016) which gathered businesses, public officials, and activists under the tagline 'advancing human-centric personal data'. Moreover, this development is supported by new regulations, including the updated EU General Data Protection Regulation (GDPR) (EU 2016), as indicated by the European Data Protection Supervisor (2016) in its opinion on systems for 'more user empowerment in managing and processing personal data'.

Developers of PDSs come from varying starting points. Accordingly, their services have varying practical solutions for data storage and sharing. Despite this, their visions exhibit a common belief that people should be able to exercise more control over their data, and that this would lead to valuable outcomes both for people themselves and for commercial service providers through more efficient markets for personal data. This highlights the focus of their approach: making possible what they consider desirable use of data, rather than prevention of misuse. Common beliefs also include the idea that there needs to be an intermediary service through which control of data by the user becomes possible.

Historically, PDSs can be seen as a rejuvenation of mid-1990s discussions related to market solutions to informational privacy. Noam (1995) considered a situation in which consumers restrict the distribution of their data by paying companies that have collected them, concluding this would likely be unsuccessful, because if data were really worth paying for, third parties could always outbid consumers. Laudon (1996) discussed the need for individuals to receive fair compensation for the use of their information, and suggested information to be deposited on accounts in information banks, which were to be the means for individuals to tap into regulated information markets. Hagel and Rayport (1997) envisioned

‘infomediaries’ as custodians and bargaining agents acting between consumers and businesses, making it possible for consumers to gain useful services in exchange for data, and for companies to access a broad array of consumer data. The latter two concepts were based on people claiming ownership of their data. In the pre-internet technological context, data ownership, paired with an intermediary facilitating data exchanges, was expected to shift power towards consumers.

Today, strategically positioned online companies are able to collect and make use of a broad array of data and provide services that Hagel and Rayport thought to be possible only through infomediaries. The term ‘infomediary’ has since served to signify a variety of services offered by aggregating information from many sources (Ho and Tang 2001), and for example today’s data brokers (Crain 2016; US Senate 2013) can be considered specific kinds of infomediaries. However, the idea of a trusted data custodian acting specifically on behalf of individuals is currently reincarnated in the imaginaries of PDSs.

### *Methodological Considerations*

To analyze these social imaginaries, I concentrate on three PDS services. Two are products of startup companies (Cozy Cloud and Meeco), and the third is an outcome of research project at MIT (OpenPDS). They all aim to enable users to first store data in a personal space, and then to make use of these data by sharing them with third parties. All attempt to carve themselves an intermediary position between individuals and companies. Focusing on the individual, they attempt to induce systemic changes through participation of individuals. The kinds of data these PDSs cover span from mundane everyday data to log-type metadata.

These PDSs exemplify ‘work in progress’. They are attempting to shape the market, working in a dynamic manner towards a more robust economic field for PDSs. While the three examples likely cannot cover all potential aspects of the PDS concept, they represent variations of imaginaries of how data collection and use should work. As laid out above, these imaginaries have also wider resonance, in terms of other similar initiatives and the policy interest they have attracted.

Material on the three PDSs includes explorative interviews with their developers and their responses to a policy questionnaire collected by the European Commission as background information for a roundtable discussion (European Commission 2016). The author was later provided with access to the responses. The aim of analysis of interview transcriptions and questionnaire responses was to identify the features that potentially afford the users agency over data. Analysis was based on iterative coding, focusing on what end-users were doing, or imagined to be doing in the future, with the PDS. In the next two sections, I first describe three PDSs and then highlight the aspects of agency over data they propose for users.

## **4. Personal Data Spaces: Descriptions**

### *Personal Cloud Server*

Cozy Cloud (2017) is a ‘personal private cloud and an app platform’ developed by a French startup company. In practical terms, Cozy Cloud is a server that a user can either set up for themselves, or have a service provider set up on their behalf. Cozy Cloud developers envision users would store data that is otherwise spread amongst databases of different service providers. The data would include mundane everyday data such as photos and documents; banking and other financial data; and data produced by activity loggers and smart home devices. Users could then make use of the data by installing and running applications on the server. In Cozy Cloud, there is a sense of resistance towards established actors: they advertise the possibility to ‘ungoogle your digital life’ by ‘reclaiming crucial parts’ of it.

The benefits Cozy Cloud expects from its service are based on possibilities to combine data from multiple sources and applications from multiple providers. Cozy Cloud promises its users a ‘frictionless data experience’ within the confines of a personal server. They also propose ‘breaking the proprietary silos’ of

data holders: Cozy Cloud expects users to employ their rights to download and store data from one company (for example, a smart thermostat provider) and then allow an application provided by another company (in the example, an electricity company) to access them. What data applications can access, and whether something is communicated outside of the server, is left for the users to decide. Further promises are based on users running Cozy Cloud on the server of their own choosing, independent of specific service providers. Cozy Cloud software is distributed for free under an open source license, meaning skilled users can set up a server on their own and even create modified versions of the software.

At the time of collection of material for this article, Cozy Cloud envisioned its paying customers to be other businesses, including companies offering Cozy Cloud servers to end-users who are unwilling to maintain their own servers, or companies developing tailored applications to be installed by the users. Cozy Cloud, then, positions itself as a provider of platform technology. Cozy Cloud does not intend to monetize or access data stored by its users, a promise they also employ in marketing. However, the developers maintain that third-party applications could reach any kinds of agreements with the users, including for example using data for targeted marketing.

### *Digital Life Management*

Meeco (2017) is a 'life management platform' developed by an Australian startup company. Meeco is a cloud service, accessible via a web browser or a mobile device, in which users have accounts. Its intended use is to create, manage and share datasets. Its marketing material has a clear privacy focus: Meeco promises to have no knowledge of the data that is stored within its service; it promises encryption of stored data; it advertises to 'never sell your data and sharing is always on your explicit terms'; and it provides communication and web browsing functions with promises of privacy and tracking-free service.

Meeco envisions its users would create datasets on concrete objects such as 'house' or 'car' or more abstract things like health, finances or plans. The contents of these datasets might include documents, measurement results, characteristics, preferences, or connections to other things. Users would share selected datasets with third parties, such as service providers, for specific purposes and on their 'own terms and conditions': the users would, for example, share certain health-related datasets with their doctor, or datasets related to purchase intentions with potential vendors. Users would individually judge what data sharing is beneficial, and under what terms. Meeco, then, intends that users explicitly exchange data for things they value.

Meeco promotes its users as both accumulators of nuanced data about themselves and as sources of abstract kinds of data such as preferences or intentions. The latter bear resemblance to the prediction products that online businesses currently produce based on extracted raw data. Indeed, Meeco views the current practices of data online tracking, data brokering and data analysis as leading to low-quality data about people. Meeco's value proposition for data-using services is that data about preferences and intentions acquired directly from people would be more accurate. Meeco, then, envisions data exchanges between users and businesses that lead to increased value for both parties of the exchange.

The details of Meeco's own business model were not determined at the time of collection of material for this article, reflecting its work-in-progress nature. Its plans were in one way or another related to monetizing data exchange markets between users and service providers: for example, based on transaction fees, subscriptions for tailored exchange services, or licensing software to run specific kinds of data exchanges.

### *Personal Data Store*

OpenPDS (de Montjoye et al. 2014; OpenPDS 2017) is a 'personal metadata management framework' developed in an MIT research project. It focuses on automatically generated log-type behavioral data from sensors, credit card transactions, or use of devices. OpenPDS developers argue that people currently do

not receive the best possible services with their data. According to them, more data would lead to better analytics and better services, but it is difficult for people to both practically provide data and to ensure that privacy is preserved in the course of data analysis. The purpose of OpenPDS is to solve this dual problem.

To make the practical provision of data possible, openPDS provides a space for its users to accumulate metadata over time. Users would, for example, share their location data to third parties by giving access to location data stored in openPDS instead of the sensor output of their mobile device. By gaining access to data through openPDS, service providers could potentially access historical data, or data from multiple data sources. The rationale is that users could decide for themselves if a service provides enough value, taking into account the data it asks to have.

OpenPDS approaches the privacy preservation target by providing access to behavioral data stored in openPDS in a way that would prevent re-identification and unsolicited further use of data. Processing of sensitive data would happen within openPDS, and only results considered non-sensitive would be sent outside. In practice, requests for data would be sent to openPDS in the form of ‘questions’. Answers to these questions, instead of the original data, would be sent back. In the example of location data, instead of sending raw historical coordinates to a service provider, the system could provide answers to questions such as ‘has the user visited X?’

At the time of material collection, the openPDS project and the development of the service itself seemed to have stalled. But, even if openPDS remained a research project without attempts to entice end-users in the longer term, it offers a relevant complementary view compared to the commercial PDSs.

The fates of these three PDSs currently remain undetermined. Individual examples of an emerging service type might well not succeed in the longer term, and the work-in-progress nature of PDSs also means that details of features may change quickly. With this in mind, instead of focusing more deeply on the specificities of these three PDSs, I turn to analyze the ways they envision users to act towards data. The purpose of this is to abstract the analysis from the features of individual PDSs, approaching them as examples representing underlying ideas about how the digital economy should work. Even if these particular PDSs fail, variations of the social imaginaries they represent continue to underlie the efforts of other technology developers discontented in the current situation.

## 5. Aspects of Proposed Agency

Based on their features and development rationales, the above PDSs propose to provide users with new forms of agency over data. In this section, I highlight four aspects of this proposed agency and contrast them to the role users play in value creation in the surveillance capitalism model. The purpose of *collecting data* is not only to store data but also to allow doing things with data. The capability to act towards stored data leads to possibilities of *data intermediation* and the consequent *controlling of analytics*. Making it possible to store data on abstract things, instead of raw data, leads to the possibility of *signaling subjective data*.

### *Collecting Data*

The intention to collect data and accumulate them in personal repositories were exemplified, albeit in different forms, by all three PDSs. Data would be uploaded or input by users themselves, collected by sensors or devices and automatically stored in PDS, or transferred from other services. Transferring existing data from other services is subject to the initial data collector providing the data, which likely depends on regulatory intervention and enforcement. The data portability rights provided in the GDPR (EU 2016), for example, work in this direction, granting users rights to download data from service providers in machine-readable format.



When they accumulate data within PDS, users become active participants in the process of data collection. Data are stored in PDS through acts of inclusion, exclusion, and moderation: users choose what sources of data are to be included or left out and what pieces of data are desired or unwanted. Data in the PDS could also be accessed over time, promoting personal history making and archiving, and resulting in a feedback loop that can allow reconsidering decisions. Decisions on the contents of the data space, then, are ongoing negotiations rather than one-off decisions, and users are provided with means to continuously participate in these negotiations. In contrast to the users' role in the data extraction model of surveillance capitalism, data collection becomes subject to reciprocities and feedback loops.

### *Intermediating Data*

PDSs do not only expect users to accumulate data, but also to provide third parties with access to data. Users, then, would use the PDS to intermediate data between the initial data sources and third parties. This role of a data intermediary was particularly pertinent to the intended uses of Cozy Cloud, as demonstrated by the smart thermostat example above, but it featured in the others as well. In the value creation model of surveillance capitalism, the production of predictions begins with the extraction of data about users over time. Data intermediation by the users would alter the first part of this process: production would begin with accessing data already accumulated in the PDS. Here, the user is envisioned to act as a gatekeeper between data and third parties. Intermediation of data between services is obviously dependent on the users' ability to access collected data, which is likely dependent on data accessibility rulings of privacy regulators.

All three PDSs envisioned users would allow third parties to access data only when it is associated with sufficient benefits. They believe users would actively seek new uses for existing data, with the expectation that this would open up currently inaccessible data to new service providers.

An element of intermediation is the terms under which it happens. PDSs maintain that decision rights for data stored in the PDS would remain with the user, and companies would be able to access data only for purposes specified by users. The PDS users, and not initial data collectors or other companies, would determine who gets to use data. This contrasts with the tendency of companies to accumulate decision rights in surveillance capitalism. Features intended to allow users to specify terms for data use included temporal or purpose restrictions on data uses, the possibility to modify data before sharing them, and the possibility to withdraw previously shared data.

### *Controlling Analytics*

A further possibility for acting towards data proposed by PDSs is to control analytics run on the data. PDSs propose two means to gain access to data analytics capabilities: users can run analytics within the PDS, or share data with providers of analytics services. With Cozy Cloud, for example, users can install data analytics applications within the PDS. With openPDS, data are processed within the service first, and only after this can they reach the value creation processes of businesses. The production of predictions by businesses, then, would not be based on the analysis of raw data; instead, openPDS performs data analysis on behalf of its user, turning raw data into something resembling intermediate products.

The proposed control not only means choosing desired analytics, but extends also to preventing undesired ones. The purpose of performing analytics within the PDS is to limit undesired uses of data, and preprocessing raw data before sending them outside works towards the same end. These features further emphasize the goal of keeping decision rights concerning data with the user. By controlling analytics, users would become participants in the production of predictions based on data. PDSs expect users would wield significant power as participants: choosing what data are used, what analytics are run, and for what purpose predictions are produced. Instead of businesses deciding how the data are used in value creation, the users would decide, based on expected benefits.

### *Signaling Subjective Data*

The above aspects of agency concern users' roles in the process of producing predictions from data. In addition to collection and sharing of data to base later analysis on, PDSs include features allowing users to share subjective data; that is, data that are timely and relevant from the personal point of view of the user. This is highlighted by Meeco, which aims to provide its users the ability to accumulate and share data on, for instance, preferences or intentions. Notably, in the value creation model of surveillance capitalism, one purpose of the production chain starting from extracted raw data is to predict such things. Meeco's aim of making users share subjective data is to increase the accuracy and quality of data that things like recommendations or personalization are based on. Online companies currently aim to increase the quality of predictions by increasing the scale and scope of data extraction, and having users share subjectively accurate data represents an alternative means towards the same end. It also represents an intervention in the production chain from extraction of raw data through data analysis to predictions. Signaling subjective data by the users would circumvent the extraction of raw data and data analysis and arrive directly at data that could fulfill the role of predictions.

## **6. A New Economic Role for Users**

Based on the above analysis, it is clear that the PDSs do not target commercial uses of data as such. They posit datafication as given and operate with the firm understanding that it leads to desirable outcomes. PDSs recognize the current benefits users receive from datafication through free services or features such as recommendations, and assume that there are more benefits to be gained. Part of their explicit aim is to increase the quality and intimacy of data, in order to achieve more detailed personalization and more accurate targeting. In this respect, their features and development rationales exemplify attempts to intensify datafication. It is, then, clear that PDSs do not aim to counter the monetization of data that lies at the heart of surveillance capitalism.

However, it is also clear that these PDSs are born from a certain discontent with how this monetization currently happens. The predictions of surveillance capitalism are monetized in markets oriented to serve advertisers and other businesses. Decisions about how users benefit from their data, then, are currently made in the context of markets facing businesses and serve the interest of platform companies that operate these markets. PDSs posit that users cannot reap enough of the current benefits, and this is where they attempt to intervene in the value creation of surveillance capitalism. They propose reorienting markets in order to change who benefits from datafication. Markets shaped by PDSs are to be consumer driven, and the user needs to decide how and under which terms data are used. In contrast to the transparency initiatives critically analyzed by Crain (2016), consumer empowerment sought by PDSs is based on changing the structure of data flows in the data industry. They aim not only at transparency of data flows but at changing who gets to decide about them.

These PDSs, then, attempt to intervene in surveillance capitalism by allowing users a new role in value creation. While users remain the sources of personal data that keep the online economy running, they are to be made sources of data in an altered sense; not only objects of data extraction, but also suppliers of data. In the imaginaries of PDSs, users have new roles in different points of the value chain: they supply raw data and intermediate products, or even final products in the form of subjective data. At the same time, PDSs underline the need for privacy-consciousness and features that allow users to limit the ways and purposes data are used. So, while the quality or quantity of data that businesses can access is expected to increase, so too is the ability of users to exercise control over the uses of data. The proposed role for individuals, then, is a data-supplying and benefit-demanding participant of the data economy.

The economic role these PDSs envision for themselves is a neutral platform provider that facilitates different kinds of data exchanges. Their business models aim at monetizing either the provision of platform technology or data transactions, but not the data itself. This is reflected also in the promises made

by the PDS developers to not even know what data are stored by users. While these services do not aim to monetize data of users, third parties could do it if it was agreeable for the users. Monetization of data, then, is not to be shunned as such, but the platform facilitating markets would not do it directly. In this sense, too, PDSs attempt to intervene in, but not counter, surveillance capitalism.

PDSs work with the assumption that new opportunities for valuable services emerge when users become data suppliers. Market mechanisms, then, are assumed to ensure new services are designed for users to choose from. These services would be provided by consumer-facing equivalents of the ‘analytical deputies’ (Degli Esposti 2014), that specialize in providing services for customers who lack analytical capabilities themselves. In other words, user empowerment is expected to happen through market mechanisms.

It is also through these analytical deputies that PDSs seek systemic changes. First, they propose building environments that offer alternative routes to market success. Surveillance capitalism favors the ability to accumulate proprietary data assets, and the effectiveness and value of predictions are increased by leveraging the scope of data collection. Therefore, businesses that do not have access to data assets are in a disadvantaged position. PDSs attempt to intervene by providing an alternative path to market success: companies could thrive by promising valuable services and analytics based on consumer-provided data. Second, in an environment where users act as data intermediaries, businesses would lack monopolistic control of data and analytics. Even if incumbent businesses continue to accumulate data, they would not be capable of unilateral market control. In this sense, PDSs resemble Baack’s (2015) open data activists. Like the activists, PDSs work to reorient datafication and break the access and interpretation monopoly that institutional data collectors have on data. Likewise, PDSs similarly recognize the need for, and aim to act as, data intermediaries for the purpose of realizing these outcomes.

To summarize the imaginaries of PDSs, they aim to intensify datafication but promise more individual control over its outcomes. Their assumption is that technical features to control data lead to the ability to control data analysis and, therefore, behavioral predictions. This hinges on premises that deserve to be spelled out; if they do not hold, even more, and more nuanced, user data ends up being produced and collected for the purposes of knowing, controlling and modifying behavior.

To begin with, PDSs expect users to peruse market offerings for desirable uses for data they have accumulated and exchange their data for things like insights, personalization, and better services. The users’ data, then, effectively turns into an object of exchange. PDSs work with a particular notion of users as subjects and data as an object: users need to consider their data as a resource to tap into, and utilize it in a way that works to their advantage. This means users must be interested and capable of taking part in managing data and also need to accept the consequences of their decisions. Given the personal nature of PDSs, these consequences are implicitly assumed to be individual in nature. By assumption, data stored in a PDS is a personal resource that should be controllable by the individual, for subjective and private benefit.

Considering data as a resource for users means they would need to make informed decisions about the use of this resource. The provision of consent for data extraction is based on a similar idea: that users weigh the costs and benefits of data collection and use in each case (Solove 2013). Solove argues that an individual’s ability to perform cost-benefit analyses on data is limited by the available information and the bounds of rational decision-making. The production of predictions based on disclosed data (Mai 2016) further complicates the issue, as the costs and benefits depend on other data and on data analysis technologies available to companies now and in the future. Notably, Crain (2016) argues that the data industry is structurally incompatible with the possibility of people being informed about data use due to trade secrets, complex market arrangements, and analytic processes that obscure the sources and

destinations of data. These issues remain pertinent: the problems of informed consent will likely hinder meaningful decision-making on data in the context of PDSs as well.

Finally, for control to be meaningful, the technical features to control data would need to result in freely made decisions. Zuboff (2015) considers power online to be identified with the ownership of means of behavior modification, and Yeung (2016) highlights the potent and unobtrusive ways that behavioral predictions produced by big data techniques are used to modify behavior. Even if PDS users' decisions were informed, the issue of behavior modification remains. Personalized choice environments enable nudging that can predictably direct people towards preferred choices without forcefully limiting the choices available (Yeung 2016). Despite increased agency towards data promoted by PDSs, businesses may remain in a position to affect—through, for example, nudging—the decisions users make.

## 7. Conclusions

The proposed agency over data reflects PDSs' efforts to reshape the economic role of users, turning them from passive data sources and objects of surveillance into active subjects and participants in value creation. In the imaginaries of PDSs, the individual is the beneficiary of datafication and the final arbiter of value derived from data. Zuboff's (2015) questions about data and learning—who can learn from data, and who decides—are, then, answered with 'individuals decide for themselves based on subjective benefits'. The success of this hinges on the possibility to freely and meaningfully carry out these decisions. If PDSs provide users with efficient means to make personal data available, but the underlying power imbalances remain untouched, they risk turning people into helpful accomplices for more efficient ways of commodification and monetization of data.

To move from the margins of the digital economy, PDSs must reorient the current institutionalized market model of surveillance capitalism. Likely, this will not happen without the aid of supportive legislation. The regulatory environment in the EU seems to take steps in a favorable direction with the data portability rulings of GDPR. They provide individuals with new rights to access data about themselves in a machine-readable format. This could make the movement of data between service providers possible and operate in favor of intermediaries that promise valuable uses for these data. The potential to use PDSs to avoid state surveillance online might well pull in the opposite direction, to the extent that state surveillance is made possible by the practices of the incumbent companies. The exploration of PDS-related concepts by incumbents (Gurevich et al. 2016) also emphasizes their capability to adapt to new regulations and to occupy new positions opened up by societal developments, which should not be underestimated either.

The eventual success of PDSs is an open question, but given the signs of traction the underlying idea has gained, we will likely see more efforts towards their development in the near future. This calls for discussion of aspects that will likely affect the success of their imaginaries of agency over data.

PDSs operate on an individual scale and build on assumptions about the kinds of needs and wishes people have. Success of PDSs depends on a large enough proportion of people aligning with these assumptions. In part, this is a question of evolution of attitudes towards surveillance: can the resigned cynicism and rationalization of surveillance (Zuboff 2015) and the feelings of powerlessness to contest industry practices (Andrejevic 2014) be turned into a strong enough social demand for alternatives? One indicator that such evolution could be taking place are the above-mentioned regulatory developments.

Apart from individual desires, PDSs also rely strongly on individuals' ability to manage data if technical means are provided. More critical questions can be posed regarding this ability. The role PDSs promote for people is demanding—is it reasonable to expect people to be willing or knowledgeable enough to control data? Even if people had control, how can it be turned into meaningful choices? In what sense does control of data lead to control of prediction, let alone modification, of behavior? On a more systemic

level, the individual empowerment that PDSs promote relies on the effect that technical means to control data have on the wider environment. To be effective in offering an alternative to surveillance capitalism, control of personal data needs to be the determinant of power. Is it reasonable to assume that once data are available, markets provide analytical capabilities for the benefit of people? How can we ensure that the ability to control data flows does not place people under more effective surveillance, in a position where they are forced to share even more details of their personal life?

This does not mean that providing people with more control would be flawed as a concept, but rather that technical features to control are not enough. The above questions regarding individual abilities and systemic effects are fundamentally tied together in that both can be supported by governance mechanisms, which are also works-in-progress in this early stage of PDSs. Control could, for example, be coupled with boundaries on how it can be exploited by businesses. By exploring alternatives to limit actions of both individuals and businesses, we could start finding mechanisms to encourage societally desirable outcomes and to ensure that the power to make decisions does not actually slip back to businesses.

## Acknowledgments

I would like to thank the Editors and anonymous reviewers of *Surveillance & Society* for their insightful comments, Minna Ruckenstein for the extremely helpful feedback, and Jogi Poikola and Teppo Valtonen for their help with material collection. Research was funded by Tekes, Grant no. 2676/31/2015.

## References

- Abiteboul, Serge, Benjamin André, and Daniel Kaplan. 2015. Managing Your Digital Life. *Communications of the ACM* 58 (5): 32–35. doi:10.1145/2670528.
- Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. 2015. Privacy and Human Behavior in the Age of Information. *Science* 347 (6221): 1–4. doi:10.1126/science.aaa1465.
- Andrejevic, Mark. 2014. The Big Data Divide. *International Journal of Communication* 8: 1673–89. Accessed July 11, 2017. <http://ijoc.org/index.php/ijoc/article/view/2161>.
- Andrejevic, Mark, and Kelly Gates. 2014. Big Data Surveillance: Introduction. *Surveillance & Society* 12 (2): 185–96. Accessed July 11, 2017. [http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/bds\\_ed](http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/bds_ed).
- Baack, Stefan. 2015. Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism. *Big Data & Society* 2 (2). doi:10.1177/2053951715594634.
- Callon, Michel, and Fabian Muniesa. 2005. Peripheral Vision: Economic Markets as Calculative Collective Devices. *Organization Studies* 26 (8): 1229–50. doi:10.1177/0170840605056393.
- Castells, Manuel. 1996. *The Rise of the Network Society. Vol. 1 of The Information Age: Economy, Society, and Culture*. Cambridge: Blackwell Publishing.
- Chignard, Simon. 2013. A Brief History of Open Data. *Paris Innovation Review*. Accessed July 11, 2017. <http://parisinnovationreview.com/2013/03/29/brief-history-open-data/>.
- Clarke, Roger. 1988. Information Technology and Dataveillance. *Communications of the ACM* 31 (5): 498–512. doi:10.1145/42411.42413.
- Cozy Cloud. 2017. Cozy Cloud website. Accessed July 11, 2017. <http://cozy.io>.
- Crain, Matthew. 2016. The Limits of Transparency: Data Brokers and Commodification. *New Media & Society*. doi:10.1177/1461444816657096.
- de Montjoye, Yves-Alexandre, Erez Shmueli, Samuel S Wang, and Alex Sandy Pentland. 2014. openPDS: Protecting the Privacy of Metadata through SafeAnswers. *PLoS One* 9 (7). doi:10.1371/journal.pone.0098790.
- Degli Esposti, Sara. 2014. When Big Data Meets Dataveillance: The Hidden Side of Analytics. *Surveillance and Society* 12 (2): 209–25. Accessed July 11, 2017. <http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/analytics>.
- EU. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data. *Official Journal of the European Union* L119/1.
- European Commission. 2016. An Emerging Offer of ‘Personal Information Management Services’ - Current State of Service Offers and Challenges. *European Commission Report*. Accessed July 11, 2017. [http://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=40118](http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=40118).
- European Data Protection Supervisor. 2016. EDPS Opinion on Personal Information Management Systems. Towards More User Empowerment in Managing and Processing Personal Data. *Opinion 9 / 2016*. Accessed July 11, 2017. [https://edps.europa.eu/data-protection/our-work/publications/opinions/personal-information-management-systems\\_en](https://edps.europa.eu/data-protection/our-work/publications/opinions/personal-information-management-systems_en).
- Evans, David S, and Richard Schmalensee. 2011. The Industrial Organization of Markets with Two-Sided Platforms. In *Platform Economics: Essays on Multi-Sided Businesses*, edited by David S. Evans, 2–29. Competition Policy International.

- Accessed July 11, 2017. <http://ssrn.com/abstract=1974020>.
- Gawer, Annabelle. 2014. Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework. *Research Policy* 43 (7): 1239–49. doi:10.1016/j.respol.2014.03.006.
- Gerlitz, Carolin, and Anne Helmond. 2013. The like Economy: Social Buttons and the Data-Intensive Web. *New Media & Society* 15 (8): 1348–65. doi:10.1177/1461444812472322.
- Gillespie, Tarleton. 2010. The Politics of ‘Platforms.’ *New Media & Society* 12 (3): 347–64. doi:10.1177/1461444809342738.
- Gurevich, Yuri, Efim Hudis, and Jeannette M. Wing. 2016. Inverse Privacy. *Communications of the ACM* 59 (7): 38–42. doi:10.1145/2838730.
- Hagel, John, and Jeffrey Rayport. 1997. The Coming Battle for Consumer Information. *Harvard Business Review* 75: 53–65.
- Helmond, Anne. 2015. The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society* 1 (2). doi:10.1177/2056305115603080.
- Ho, Jinwon and Tang, Rong. 2001. Towards an Optimal Resolution to Information Overload: An Infomediary Approach. *GROUP '01 Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*: 91–96. doi:10.1145/500286.500302.
- Kelty, Christopher M. 2008. *Two bits: The Cultural Significance of Free Software*. Durham: Duke University Press.
- Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.
- Laudon, Kenneth C. 1996. Markets and Privacy. *Communications of the ACM* 39(2): 92–104. doi:10.1145/234215.234476.
- Mai, Jens-Erik. 2016. Big Data Privacy: The Datafication of Personal Information. *Information Society* 32 (3): 192–99. doi:10.1080/01972243.2016.1153010.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data. A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.
- Meeco. 2017. Meeco website. Accessed July 11, 2017. <http://meeco.me>.
- MyData. 2016. MyData 2016 Conference website. Accessed July 11, 2017. <http://www.mydata2016.org>.
- Noam, Eli M. 1995. Privacy in Telecommunications: Markets, Rights, and Regulations. Part III: Markets in Privacy. *New Telecom Quarterly* 4Q95: 51–60.
- OpenPDS. 2017. OpenPDS website. Accessed July 11, 2017. <http://openpds.media.mit.edu>.
- Poikola, Antti, Kai Kuikkaniemi, and Harri Honko. 2015. *MyData – A Nordic Model for Human-Centered Personal Data Management and Processing*. Finnish Ministry of Transport and Communications. Accessed July 11, 2017. <http://urn.fi/URN:ISBN:978-952-243-455-5>.
- Richards, Neil M., and Jonathan H. King. 2014. Big Data Ethics. *Wake Forest Law Review* 29. Accessed July 11, 2017. <https://ssrn.com/abstract=2384174>.
- Rieder, Bernhard, and Guillaume Sire. 2013. Conflicts of Interest and Incentives to Bias: A Microeconomic Critique of Google’s Tangled Position on the Web. *New Media & Society* 16 (2): 195–211. doi:10.1177/1461444813481195.
- Rochet, Jean-Charles, and Jean Tirole. 2003. Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* 1 (4): 990–1029. doi:10.1162/154247603322493212.
- Solove, Daniel J. 2013. Privacy Self-Management and the Consent Dilemma. *Harvard Law Review* 126 (7): 1880–1903. Accessed July 11, 2017. <https://ssrn.com/abstract=2171018>.
- Spiekermann, Sarah, Alessandro Acquisti, Rainer Böhme, and Kai-Lung Hui. 2015. The Challenges of Personal Data Markets and Privacy. *Electronic Markets* 25 (2): 161–67. doi:10.1007/s12525-015-0191-0.
- Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge: Polity Press.
- Taylor, Charles. 2004. *Modern Social Imaginaries*. Durham, NC: Duke University Press.
- U.S. Senate. 2013. *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes*. U.S. Senate Committee on Commerce Science and Transportation. Accessed July 11, 2017. <http://1.usa.gov/1vIVESn>.
- van Dijck, José. 2014. Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance & Society* 12 (2): 197–208. Accessed July 11, 2017. <http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/datafication>.
- Yeung, Karen. 2016. ‘Hypermudge’: Big Data as a Mode of Regulation by Design. *Information, Communication & Society* 20 (1): 1–19. doi:10.1080/1369118X.2016.1186713.
- Zuboff, Shoshana. 1985. Automate / Informate : The Two Faces of Intelligent Technology. *Organizational Dynamics* 14 (2):5–18. doi:10.1016/0090-2616(85)90033-6.
- Zuboff, Shoshana. 2015. Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology* 30: 75–89. doi:10.1057/jit.2015.5.
- Zuboff, Shoshana. 2016. Google as a Fortune Teller. The Secrets of Surveillance Capitalism. *Frankfurter Allgemeine*, March 5. Accessed July 11, 2017. <http://www.faz.net/-gsf-8eaf4>.

**Lehtiniemi, T. & Haapoja, J. (2020)**  
**Data agency at stake: MyData activism and alternative  
frames of equal participation**  
*New Media & Society*, 22(1), 87–104









# Data agency at stake: MyData activism and alternative frames of equal participation

new media & society  
2020, Vol. 22(1) 87–104  
© The Author(s) 2019



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/1461444819861955  
[journals.sagepub.com/home/nms](https://journals.sagepub.com/home/nms)



Tuukka Lehtiniemi  and Jesse Haapoja 

University of Helsinki, Finland; Aalto University, Finland

## Abstract

Data activism has emerged as a response to asymmetries in how data and the means of knowledge production are distributed. This article examines MyData, a data activism initiative developing principles for a new technical and commercial ecosystem in which individuals control the use of personal data. Analyzing material collected at a formative event shaping MyData activism, we examine how more just data arrangements are framed to enhance equal participation. Our analysis shows agreement on what is ultimately at stake: individual data agency and fair competition in the data economy. However, two alternatives are offered for what participation involves. Collaboration with commercial actors favors framing participation as agency in data markets, thereby potentially limiting the scope of what is at stake. The alternative framing presents a rights-based understanding of economic and civic agency, potentially leading to a broader understanding of participation in a datafied society.

## Keywords

Data activism, data agency, data economy, frame analysis, justice, MyData, participation

## Introduction

Routine aspects of our lives today produce data, which play an increasingly important role in contemporary capitalism. Companies have long stressed the social benefits, democratic potential, and consumer empowerment accruing from the collection and

---

### Corresponding author:

Tuukka Lehtiniemi, Centre for Consumer Society Research, University of Helsinki, P.O. Box 24, 00014 Helsinki, Finland.

Email: [tuukka.lehtiniemi@iki.fi](mailto:tuukka.lehtiniemi@iki.fi)

exploitation of user data (West, 2019). In contrast, scholarship forming what could be called the data economy's "counternarrative" (Pasquale, 2017) has focused on the asymmetric distribution among companies and individuals of the means of data-based knowledge production (Andrejevic, 2014; Citron and Pasquale, 2014; Crain, 2018; Tufekci, 2014; Van Dijck, 2014; West, 2019; Zuboff, 2015). Data practices that dominate the digital environment have developed alongside technologies that convert aspects of social life into quantifiable data, and ahead of ethical scrutiny, public understanding, and regulation (Zuboff, 2015). These developments have given rise to advocacy and experimentation connected with people's rights, capabilities, and roles as users, consumers, and citizens in the information society—or digital citizenship (Hintz et al., 2017). Examples include advocating the rights of consumers to participate in content production (Postigo, 2012); employing open source principles for digital rights campaigning (Breindl, 2013) or open data promotion (Baack, 2015); hacking as a form of data agency (Pybus et al., 2015); using digital media for political causes (Kaun and Uldam, 2017); the deployment of infrastructure and tools by civic hackers (Schrock, 2016); and alternative data collection and analytics practices in the Quantified Self (QS) community (Sharon and Zandbergen, 2017). Technology and advocacy movements indicate different ways of responding to the closing off and monopolizing of knowledge production and value creation in digital environments, and emerging movements may either support or resist the dominant political economy of data. The QS community, for example, engages in "soft resistance" (Nafus and Sherman, 2014) to dominant data practices by welcoming big data actors but questioning who gets to aggregate data and how. QS is also ambiguous in terms of its valuations, allowing the values of sharing to thrive alongside the commercialization of self-tracking (Barta and Neff, 2016).

This article contributes to research on *data activism*, referring to civic engagement and political action responding to the uneven distribution of data access and capabilities in datafied times (Baack, 2015; Milan and Van der Velden, 2016). Data activism "seeks to challenge existing data power relations and to mobilize data in order to enhance social justice" (Kennedy, 2018: 18), recognizing that more just practices can be promoted in the place of dominant ones (Dencik, 2018). As data activism is rooted in data and software, it can involve the promotion of alternative technologies and associated policies, which may in turn involve some form of collaboration with the industry (Milan and Van der Velden, 2016), such as the producers of data-related technologies. This collaboration can serve pragmatic ends; while technology-oriented activism requires the development and production of alternative technologies, firms producing such technologies can in turn seek markets for their outputs (Hess, 2005). In mobilizing more just data arrangements—how organizations collect and use data, the policies that govern such practices, and new capabilities for people to engage with data (Kennedy, 2018)—data activism may then concern firms as participants and beneficiaries. Our contribution is to examine the tensions that emerge between activist and commercial interests, when commercial actors are involved in data activism.

In scholarship on data justice (Dencik et al., 2016; Taylor, 2017), social harms resulting from dominant data practices are seen to both exacerbate existing injustices and produce new ones. We take the normative view that justice requires arrangements that permit all to participate as peers in social life (Cinnamon, 2017; Fraser, 2008). From this

point of view, the more just data arrangements envisaged by data activists pertain to enhancing citizen participation in the information society by removing obstacles that hamper equal engagement. Starting from this understanding of justice enables alternative views of the issues relevant to it: injustices can concern the economic dimension of distribution, sociocultural recognition, or political representation (Fraser, 2008). Dominant data practices can be seen to pose threats to equal participation in all three dimensions (Cinnamon, 2017). Asymmetric data accumulation practices give rise to distributive injustice, denying some the resources necessary for participation. This also lays the foundation for sociocultural misrecognition through profiling and social sorting, and for political misrepresentation by restricting people's means of contesting how they are represented by data. The asymmetric distribution of data can, therefore, be seen as the initial injustice that enables further injustices.

This article focuses on how equal participation is framed in data activism involving commercial actors and interests. What injustices hamper equal participation, what are their remedies, and whose interests deserve consideration? Our empirical context is MyData, a data governance initiative that originated within open data activism in Finland, and has since expanded into an international movement. MyData proponents argue that, to realize their individual, commercial, and societal benefits fully, personal data should be released from the confines of monopolistic data holders, provided that individuals "have an easy way to see where data [...] goes, specify who can use it, and alter these decisions over time" (<https://mydata.org/what-we-want>). MyData envisions a technological and commercial ecosystem where people would control the sharing of their data between interoperable data sources and endpoints. Commercial actors would occupy positions in the ecosystem as, for example, technology providers, service developers, or intermediaries. Ultimately, the expansion of this ecosystem is expected to transform individuals into "empowered actors, not passive targets, in the management of their personal lives both online and offline" (Poikola et al., 2015: 2). Even though MyData aims to increase people's capabilities to use their data, it also promises to serve firms' prevailing economic interests in personal data: "[MyData] combines digital human rights and industry need to have access to data" (Poikola et al., 2015: 4). It, therefore, provides an example of data activism explicitly involving commercial data use, making it highly relevant to our research interest.

Our data were collected at the first large international gathering of people interested in MyData's aims, which turned out to be a formative event for the MyData community. Applying frame analysis to keynote presentations and audience responses at this influential event, we examine how injustices and their remedies are presented in MyData. Our analysis identifies agreement on what is ultimately at stake: individual data agency in the information society. Dominant contemporary data arrangements are framed as hampering equal participation, the remedy being the development of a technological infrastructure providing people with agency over their data and allowing their participation in data collection, sharing, and processing. This was simultaneously framed as a means to redistribute data so that firms can equally compete in an environment currently dominated by monopolistic data holders. However, while general consensus was reached on these means of achieving equal participation, alternative framings for participation itself were suggested. One framing equated participation with the ability to choose between

alternative data uses in the market, while another considered participation more broadly in terms of rights and digital citizenship. These frames evidence multiple interpretations of specific dimensions of justice by either construing individuals as market agents or, alternatively, also allowing the consideration of economic and civic agency in broader terms. It is here, we argue, that the involvement of commercial interests in data activism becomes significant. When data agency must serve both activist and commercial interests, and market agency is more readily transformed to serve commercial data uses, what is at stake risks being reduced to participation in data markets.

## Data activism and equal participation

Data activism includes variable forms of engagement with existing data arrangements and their politics, and different ways of mobilizing more just data arrangements. By taking an unjust distribution of data as the initial injustice preventing equal participation in the information society (Cinnamon, 2017), we may examine how alternative data arrangements proposed by data activists aim to address this inequity.

Proactive data activism understands data as a potent force for social change, and sees active engagement with data as “a pathway to empowerment, equal participation and action” (Milan and Gutierrez, 2018: 58). This may mean employing data infrastructure for explicit advocacy goals, such as impeding environmental threats through data collection, sharing, and visualization, and the promotion of data transparency (Milan and Gutierrez, 2018). Here, addressing distributive injustice is a means to combat other injustices. Another example is open data activism, which advocates the redistribution of data, aiming to break the interpretative monopoly of governments, and to balance the unjust distribution of power and knowledge (Baack, 2015). Redistribution of data, however, does not automatically promote justice; open data exist only in relation to the political economy of data, and due to asymmetrically distributed capabilities to do with data, opening data might benefit corporations, but not citizens (Johnson, 2014). More broadly, the involvement of corporations in data activism has been objected due to concerns over potential co-optation, as well as dubious political alignments (see Schrock, 2016); for example, political processes restricting the counter-hegemonic potential of open data can instead shape it to support the marketization of public services (Bates, 2013). Data activists themselves can act as a monitorial elite enabled by open data, guarding the public against unjust data use (Schrock, 2016), and may also recognize the need for intermediaries that help to make open data more accessible to the public (Baack, 2015).

For personal data, an unjust distribution results from data industry’s dominant practices separating people from their data and enabling data accumulation by corporations (Cinnamon, 2017). Some data activists posit these data practices as threats to individual rights, and combat them with technical self-protection, such as anonymity, obfuscation, and encryption (Milan and Van der Velden, 2016). In response to an unjust data distribution, this kind of reactive data activism attempts to prevent the production of data in the first place, avoiding the potential harm as well as the benefits accruing from exchanges involving personal data. Here, seeking justice becomes a private act relying on technical skill and ability (Dencik, 2018). Some recent, more proactive instances of data activism focus on redistributing personal data, or their benefits, from firms to people. In

addition to MyData, developments include the “re-decentralization” initiative of web pioneer Berners-Lee, aiming to make personal data a resource for people (<https://solid.inrupt.com>; Andrejevic, 2014; Brooker, 2018); the proposal by another Internet pioneer, Lanier (2013), to achieve commercial symmetry between firms and users by remunerating people for personal data use; the development of software (see Lehtiniemi, 2017) and devices (Crabtree et al., 2016) to provide users with means to exercise control over data collection and use; and “smart disclosure” programs releasing machine-readable personal data from firms to consumer-citizens (Iemma, 2016). Whereas the data analytics industry promises to put organizations in charge of their data for their own advantage (Beer, 2018), these initiatives aim to do the same for individuals. On the surface, they seem to advocate economic agency for people in the information society. They can, however, be criticized on many grounds: for example, that they are excessively individual-centric and reliant on markets that do not work economically (Charitsis et al., 2018); that over-individualization can make them susceptible to private sector co-optation in the same way as the protection of privacy (Coll, 2014); and that they are driven by the judgments of the technical elite about just data practices (Kennedy, 2018). Despite this, they represent work-in-progress experimentation on what a more just data economy could look like; we, therefore, consider them as “moments where meaningful change can occur” (Schrock, 2016: 583). The following section describes our empirical approach to one such moment.

## Data and analysis method

### *The MyData conference*

The first author has closely followed MyData in Finland through participant observation in research projects since 2014 (see Lehtiniemi and Ruckenstein, 2019). Data for this research were collected in the context of participant observation at a conference called “MyData 2016” (<https://mydata2016.org>). In the previous year, the Finnish activists had published a report outlining MyData’s aims (Poikola et al., 2015) which attracted interest from like-minded activists around the globe, eventually leading to organizing the conference in collaboration with the nonprofit Open Knowledge Finland and the French think-tank FING. The event attracted an audience of 700 domain experts<sup>1</sup> with an interest in “human-centric personal data management,” including businesspeople, entrepreneurs, technologists, researchers, privacy advocates, and public sector officials. The event became a formative step for the MyData community, providing grounds for further developments: annual follow-up conferences, a declaration outlining MyData principles (<https://mydata.org/declaration/>) and the launch of an international NGO “MyData Global” in 2018, with the expressed goal of creating “a fair, sustainable, and prosperous digital society” (<https://mydata.org>).

If MyData is considered as an emerging field (Fligstein and McAdam, 2012) of data activism, the conference may be regarded an example of a field-configuring event (Lampel and Meyer, 2008); these are events which shape technologies, markets, or industries by assembling diverse interest groups, offering interaction opportunities and facilitating information exchange, and collective sensemaking. Actors in an emerging

field have the leeway to shape it to suit their own interests by inducing the cooperation of others (Fligstein, 2001), and a professional conference offers a venue for contestation between future visions, as well as an environment facilitating selection between alternatives (Garud, 2008). Indeed, “if the whole field were to be contained in a nutshell, a conference would be its most likely manifestation” (Garud, 2008: 1084). At the MyData conference, then, actors attempted to shape what MyData is “about” to suit their activist, policy, or commercial ends.

### *Data and analysis*

Our empirical approach is based on analyzing the frames constructed by the conference’s keynote presentations, and the reception of these frames by the audience. Frames in general offer a schema for highlighting aspects of a situation, functioning as modes for articulating strategy to be undertaken. Those constructed in keynotes suggested ways of understanding the current situation, identifying issues to act on, and ways of acting on them. As an analytical framework, we employ the identification of collective action frames (Benford and Snow, 2000; Snow and Benford, 1988) that diagnose the issue in need of change and who is to blame, prognose solutions and how to achieve them, and motivate collective action. By focusing on keynotes, we employ a form of purposive sampling of settings where the processes of interest are most likely to be observed (Silverman, 2006: 306–307). Three features make keynotes suitable for our purpose: first, the biases demonstrated by the choice of speakers, as the event organizers selected them with an eye toward shaping MyData (see Lampel and Meyer, 2008); second, keynote lectures concerned MyData’s means and ends generally rather than detailed issues such as technical or legal minutiae; and third, related to this, the majority of the conference public was present during the keynotes, necessitating that speakers navigate the varied interests of conference participants. Overall, we can expect keynote speakers to attempt to construct frames that resonate with their audience’s interests; however, while different keynote speakers represented different interests and backgrounds, investigating only the constructed frames risks devaluing the power relations in play. In order to take this into account, we also examine the success of framing efforts (Snow and Benford, 1988) through audience responses to them, allowing us to examine not only how injustices are framed as obstructing equal participation, and the means suggested for their removal, but also the extent of agreement on these issues.

Our material consists of two datasets. First, we transcribed video recordings<sup>2</sup> of 12 keynote talks and the follow-up Q&A sessions, totaling some 7 hours of recordings. Second, we received access to 750 anonymous messages sent by audience members using online backchannel software developed for real-time audience interactions at events (Nelimarkka et al., 2016). The software allowed people to send anonymous messages during keynote lectures, specifically prompting “comments and feedback to speakers” as well as “key lessons.” The messages were public to the conference audience. While commenting was continuously encouraged by conference hosts, strong agreement, and disagreement with issues raised may be over-represented. Nevertheless, we argue that this method of gathering audience data is fruitful as there is a low barrier to giving feedback, and immediate responses can be gathered from a wide range of participants. In addition,

we prepared field notes on our observations during the conference, which were employed as background material for this study. Using Atlas.ti, we initially identified sections from the keynote transcriptions that represented collective action frames and broadly concerned participation in the information society. We classified these sections with an open coding scheme, and iteratively reclassified them until reaching the six frames presented in the next section. We included only frames that were either widespread or contested. The audience interaction data were then employed to examine agreement and tensions arising in response to the identified frames.

To present our results, we divided the keynote speakers into five groups based on their affiliation: one conference *organizer*; *advocates* including an NGO representative and a journalist/author; *technology developers* from a start-up and a research consortium; speakers affiliated with *private sector* firms such as a telecom company, financial services companies, and a consultancy; and speakers from the *public sector* including a ministry official, a data protection authority official, and a Finnish government minister. Two of the speakers came from Finland, the others from elsewhere in Europe, Australia, and the United States. We also include quotations from anonymous *audience* members.

## Framing MyData

An overview of the frames of participation—under three headings—and how they were employed in keynotes, is presented in Table 1. *Participation enablers* exhibit a widely employed frame describing how favorable developments in technological and regulatory environments make promoting new data arrangements possible. *Agreed-on means of participation* include two frames identifying key injustices and their remedies. One was the inability of people to act on their personal data, with the solution being to develop technologies that provided users with data agency, and another was the asymmetric access to personal data that hindered firms' opportunities; both were framed to allow simultaneous redistribution of data between firms. These frames were employed by all speaker groups and widely accepted by the audience. *Contested aims of participation* include alternative framings that also received contrasting audience responses. Notably, many speakers avoided them completely. One contested issue was whether the data economy's giants should be allowed to benefit from opportunities emerging from data activism. Alternative frames were also constructed for what equal participation involved. Some speakers, including technology developers, framed equal participation as market symmetry between users and firms. The alternative was to frame participation as based on rights and citizenship.

### *Participation enablers*

*Technical and legal tools.* Two major developments were framed as enabling dominant data arrangements to be challenged: evolving personal data technologies and a changing regulatory environment. The technological driver was the increasing availability of personal data technologies for individuals to use for their own benefit. While data collection, storage and analysis had so far been only available to corporations, the underlying technologies were reaching a level of mundanity and ubiquity, which meant that individuals could

**Table 1.** Frames and keynotes.

Keynote speakers	Frames					
	Participation enablers		Agreed-on means of participation		Contested aims of participation	
	Technical & legal tools	Agency for individuals	Redistribution of data	Beneficiaries	Market symmetry	Fundamental rights
Conference organizer	x	x	x			
Activist 1	x		x			
Activist 2		x				x
Developer 1	x	x	x		x	
Developer 2	x	x	x	x	x	
Private sector 1	x	x	x			x
Private sector 2		x				
Private sector 3	x			x		
Private sector 4	x	x	x	x	x	
Public sector 1			x			
Public sector 2	x	x	x			
Public sector 3	x		x		x	



claim control of their data. This was framed as technology democratization countering unjust data arrangements: “the only thing that has been limiting us up until now, is the ability for us to have that technology” (*Developer 1*). The regulatory driver were rulings to ensure data access and interoperability, whose role was instrumental; in order for people to have control over their personal data, new services would need to work together, and data would need to be accessible and technically and semantically interoperable. Of notable importance was the then-upcoming EU’s General Data Protection Regulation containing data portability rulings ensuring machine-readable access to personal data: “GDPR is really important and is about to [...] rebalance [...] the relationships between individuals and companies” (*Public sector 2*).

The significance of these developments was not contested and their combination was an opportunity to shape new data arrangements: “[Data] portability is really a legal tool that we will be able to mobilize for the MyData projects” (*Private sector 1*). The responsibility of the MyData community was to ensure that the opportunities are properly exploited: “the legal tools [...] become really useful if they meet a social movement, if they meet a cultural change” (*Private sector 1*).

### *Agreed-on means of participation*

*Agency for individuals.* Central to diagnosing the injustice of dominant data arrangements was that they worked in the interest of firms and organizations, but not individuals. The majority of keynote speakers mentioned the inability of individuals to act in relation to data, presenting the digital environment as detrimental to human agency. As individuals did not have the meaningful capability to make decisions on their data, their choices were constrained: “The idea of [...] a complete opt out or this total surveillance [is] no agency at all. That’s not a social contract that’s sustainable” (*Developer 1*). The culprit was technology that only firms could use for their own benefit: “Why is it we cannot have more freedom to do digital stuff ourselves? Because we don’t have our own platform” (*Developer 2*).

The prognostic component of this frame was the development of technologies allowing the control of data use, transforming individuals from objects of data collection into subjects with data agency. MyData was about “empowering people with their data” (*Organizer*) or “engaging with information in a way that actually enriches our life” (*Developer 1*). Agency was hence framed as the capacity to decide who can use data and on what terms. “Personal agency systems” (*Developer 1*) would allow people to use data for their own advantage; selective sharing of data would enable the conveyance of abstract notions such as intentions or preferences, leading to the fulfillment of personalized wants and punctual service delivery. Agency was framed as the defining feature of MyData: “PIMS<sup>3</sup> are when you give individuals agency through new technologies [...]. It’s about something that’s personal and mine, [that] understands me [...] and acts in my interest” (*Private sector 4*). In addition to individual benefits, agency was framed as realizing societal benefits; the developed technologies would lead to an information society characterized by individual rights and free will, in which people would participate by making their voices heard.

A concern was that what was being offered would not be recognized: “People have been formatted for 20 years to get excellent services without caring about their data”

(*Private sector 1*). The problematic assumption of willing and capable technology users was, however, to be addressed by augmenting human capabilities with technology: “We don’t want to be constantly processing [...] our consent [rather] we will be able to outsource some of these things” (*Developer 1*). Audience reactions were largely supportive of the identified problems and proposed solutions; for example, as one person observed, “We have no agency today. But can’t we build it back? Via MyData tools?” (*Audience*). Some comments, however, were aimed at broadening the view to extend beyond technological solutions, such as “The real question is who sets the norms?” (*Audience*), and, “How to make people desire that agency?” (*Audience*).

**Redistribution of data.** Equally prominent was a frame diagnosing another asymmetry in the data economy: the unjust distribution of data between firms. Its prognostic element framed the technologies providing data agency to individuals as also benefiting firms.

Data economy kingpins were presented as being successful due to how data aggregation and monopolization further cemented their position: “Gathering a lot of data is kind of [an arms race]. It’s not a game we can win, because we are a small company” (*Private sector 3*). At the same time, the economic model, based on commercial surveillance, was framed as erosive, raising doubts about its sustainability: “Trust towards organizations [...] has never been so low. And business data practices play a big part in that growing mistrust” (*Organizer*). Correcting the unjust distribution of data was required: “We have to move from winner-takes-all to competition-takes-all” (*Private sector 1*). Individual agency in relation to data was expected to bring about a “disruption to current data aggregator models” (*Public sector 1*). When individuals can decide how and by whom data are used, it will no longer be possible to build monopolistic positions on proprietary data assets; instead, people will share personal data with firms and organizations that serve their interests. Competition for users’ data would not only reinstate trust in data-using businesses in general, but would also provide a competitive edge to firms that earn consumers’ trust: “The more trusted you [are], the more data you’ll be able to handle and collect from the individual, [and] the more revenue you create” (*Private sector 4*). This would lead to opportunities “for you and I to absolutely revolutionize the creation of new value” (*Developer 1*) by means of new innovative services.

Reacting to this framing, vocal audience members demanded concrete evidence of business success, pointing out, for example, “Every time monetary values come up, discussion gets vague and disconnected from reality” (*Audience*). Converting visions to concrete reality was thought to require not only abstract promises of business opportunities, but also evidence of commercial success.

### *Contested aims of participation*

**Beneficiaries.** This frame concerned the interests that could be served by the business opportunities which, it was expected, equal participation in the data economy would create. At issue here were the dominant players—or GAFA<sup>4</sup> as they were referred to—and whether they should be strictly resisted, or whether data agency was what always mattered.

On most occasions when mention was made of GAFAs, MyData was about explicit resistance: there was “a battle to address” (*Private sector 3*). One speaker, a technology developer, first expressed the will to collaborate and share technology with “anyone who feels the way we do.” When directly asked about GAFAs by the audience, the speaker stated, however, that they “don’t want Facebook there” (*Developer 2*). Resisting GAFAs was, societally, the right thing to do: “The rules which have been laid down by GAFAs could represent a threat for liberty and [...] the free market” (*Private sector 3*). The audience humorously supported resistance: “How do we kill Google, Apple, Facebook and Amazon?” (*Audience*) and “Google and Facebook are fundamentally doomed” (*Audience*).

The other way to regard GAFAs was less explicit and inclusive of anyone following the MyData principles. Large corporations in particular would react slowly, so patience should be exercised, and inclusion in MyData should be based on future actions. An example of attempting to include GAFAs was this avoidance of drawing boundaries:

Is Facebook a PIMS? I think platforms [...] that give individuals agency [...] can start to be considered as PIMS. [...] We need to be very careful of thinking that PIMS are a binary. (*Private sector 4*)

In this view, it was not important to categorize the firms, but rather to consider whether their technologies “gave individuals agency.”

Many speakers did not express their position on this issue. This reluctance was evident to the extent that GAFAs were on more than one occasion referred to as “the elephants in the room.” The audience had no such restraint. Several anonymous audience comments, for example, directly demanded the above speaker to acknowledge a previous consultancy relationship: “Facebook has hired you [so] the goal must be to sell Facebook as a PIMS?” (*Audience*). The tension over who should be allowed to benefit from user data mainly emerged through audience responses.

*Market symmetry.* Above, individual agency with regard to data was framed as a prerequisite for participation in the information society. Data agency would transform individuals into empowered subjects; however, framing agency and participation relied on two different understandings of what participation involved, so we begin by discussing how parity of participation was framed as market symmetry.

The asymmetric relationship between individuals and firms was framed as arising from the inability of individuals to exercise economic interest in terms of their data. The problem was the asymmetric commodification of data by commercial players; consequently, this frame extended the commodification of personal data so that they would become saleable, or rather exchangeable, by individuals themselves. The offline world offered an illustrative comparison: “We have many more freedoms in physical lives [...] because [...] we have freedom of property. By owning stuff, we are free to use it make our lives better” (*Developer 2*). The objective was to shape an economy where “customer data is not just a corporate asset, but also a personal asset” (*Private sector 4*). This framing presented individuals as market agents, data agency as market agency, and participation in the information society as making choices in the marketplace from different

options for data use. Benefits on the societal level would emerge from the rational actions of individuals who were treating their data as an asset serving their own interests. As individuals seek to make their lives better by exchanging data for services, competition between firms to provide these services would ensure the best possible options from which to choose. Many reactions from the audience were supportive of individual benefit-seeking through data markets, something enunciated in the comment, “Love the idea [of] helping people achieve outcomes and experiences they desire and that have real value to them” (*Audience*).

*Fundamental rights.* The second framing of participation by means of data presented equal participation in data collection and use as something resembling a fundamental right. It was aligned with the market symmetry frame in the diagnosis of problems arising from the privileged economic relationship some firms had with data. However, agency was not framed as making market choices between alternative data uses, but as the right of individuals to determine what can be done with their data. While the argumentation was not extensively spelled out—possibly due to this frame’s being provoked as a reaction to the observed inadequacy of the market symmetry frame—extending the commodification of data was nonetheless seen as a dubious means to reduce the harm that commodification had initially caused:

You give me some information as if you’re handing me a pile of stuff. [...] It’s not what really goes on with participation. (*Activist 2*)

We should stop talking about owning our data. [...] We should anchor them to [...] fundamental rights, and [...] clearly refuse those approaches of people who want to monetize personal data in exchange for openness. (*Private sector 1*)

An approach rooted in rights would better contain the harm caused by the commodification of data by commercial actors. These “fundamental rights” concerned data agency in the sense of participating in processes that determine how and for what purposes data are used, such as democratic governance over sharing the value produced with data. In this model, individuals could also participate in the information society beyond the pursuit of economic self-interest:

We need to [emphasize] the community, the crowd, the strengths of collective action [...] Let’s put participation in this sense in the very center of the way we think about data. (*Private sector 1*)

The aim was, then, to produce subjectivities that would transform people from objects of data collection into digital citizens with rights and entitlements.

In audience responses, the market symmetry frame was challenged as well, mainly due to the complexities involved in the ownership of digital goods: “We may have different rights in data, [...] but not ownership like in property” (*Audience*). Audience reactions were, however, divided: for example, both “personal data is not property” and “personal data is property” were proposed and up-voted as important lessons learned at the closure of the conference.

## Discussion: the dimensions of equal participation

Our analysis shows that MyData proponents agree on the diagnosis that the lack of individuals' agency over personal data is the principal problem, and on proposing MyData technology as the means of resolving the problem. The agreed-on goal for MyData was to transform people into "proper modern agentic individuals" (Meyer and Jepperson, 2000) able to manage their lives on- and offline. This would be achieved through an ecosystem of personal data technologies providing people with the capability to make data serve their own interests rather than only those of commercial firms. Contestation over who participates and how (Zuboff, 2015) was, therefore, framed as a question that needs to be tackled with technology development, and specific kinds of technology were a condition for having agency in a datafied environment.

MyData proponents framed new data arrangements in terms of user empowerment, but simultaneously presented them as supporting the recovery of missed economic opportunities and as providing innovation potential for firms and society at large. While early discourse on the dominant arrangements of the data economy also focused on consumer power accruing from data gathering, it had largely masked companies' economic interests in data use (West, 2019). Here, in contrast, commercial data use is part and parcel of the envisioned realization of datafication's benefits, and the lack of commercial success stories to exemplify the economic potential of more just data arrangements was lamented. However, although commercial data use is in principle lauded if it involves data agency, the tensions involved in allowing the GAFA to enjoy MyData's commercial benefits demonstrate that the ethics of acceptable data use could be more nuanced.

Even if MyData proponents agreed on data agency and a more just distribution of data as the first steps toward settling further injustices and achieving equal participation (see Cinnamon, 2017), this agreement does not imply specific form of participation. Here, we identified two alternative frames. The first frame, participation as market symmetry, not only focuses on the economic dimension of participation in the information society, but it also involves narrowing the economic dimension down to market exchange. Equal participation, here, primarily signifies the ability to choose between alternative uses for personal data in the marketplace. It is proposed that the obstacles preventing equal participation could be dismantled by providing people with the means to exchange personal data, and the market is expected to take care of the rest. In this framing, MyData aims to transform people into consumer-participants in the information society (see Lehtiniemi, 2017), and to base participation in market agency. The routinization of data collection (Couldry and Yu, 2018) is not seen as a problem as such, but the aim is rather to subordinate it to the market, with the belief that suitable end-user technologies will allow people to exercise control in the sphere.

This frame constitutes an extension of data industry rhetoric presenting personal data as an asset to be turned into value (Sadowski, 2019)—in this case, for users themselves. The promise for firms is fair competition in markets for alternative data uses, where access to user data would be gained by supplying enticing services. The ability of firms to exploit data for competitive advantage would, then, not stem from a position in a locus of user activities which enables the monopolistic extraction of user data (Zuboff, 2015), but from the quality of their offerings. The value of personal data is primarily understood

to lie in exploiting data as a scarce resource: in the case of users, for the purposes of self-interest; in the case of firms, for competitive advantage. This “competitive value” derived from data is what already motivates the data industry (Cinnamon, 2017). Framing participation as market symmetry, then, does not fundamentally question the data industry’s dominant economic rationale, but rather aims to transform it to serve the ends of both activism and commercial actors.

The other alternative frame for participation is based on rights. Here, market symmetry is presented as a dubious means of achieving equal participation. Instead, people are to be transformed into digital citizens more broadly understood, with rights, entitlements, and the ability to participate in more democratic governance of data use (Cardullo and Kitchin, 2018; Evans, 2017). This frame allows data technologies to be considered as a means of not only correcting the initial distributive injustice, but also directly addressing other dimensions of it, such as misrecognition or misrepresentation (Cinnamon, 2017). The imagined data agency can be understood in terms of what Hintz et al. (2017) call ideal configurations of digital citizenship: “comprehensive self-determination in a datafied environment” (p. 735) made possible by an amalgamation of the necessary infrastructure, its informed use, an enabling regulation, and public knowledge. In terms of the economic dimension of participation, this frame also presents a broader view than merely market participation: the economic can be considered not only as meeting market demand but also, more generally, in terms of provisioning goods and services that meet the needs of humans (Elder-Vass, 2016: 28–29; Nelson, 1993). From this viewpoint, data activism seeking just data arrangements for equal participation would explicitly consider which arrangements allow provisioning for human needs. This would involve the inclusion of other kinds of value derived from data, in addition to the competitive value gained from data that others do not have. Value could be drawn from using data for the common good, or for serving the interests of specific communities (Lehtiniemi and Ruckenstein, 2019; Cinnamon, 2017). The roles offered to people could, therefore, be extended from consumers toward participants in a manner that is grounded in rights and the common good (Cardullo and Kitchin, 2018).

These alternative frames of participation, which are at least potentially at odds with each other, encourage consideration of the relationship between the involvement of commercial interests, and the goal setting of data activism. Examples from technology movements in symbiotic relations with the private sector, such as the free software movement, indicate that when a movement’s innovations are incorporated within industries, they are transformed to serve profitability concerns more effectively, potentially leading to conflicts within the movement (Hess, 2005). The QS community, however, provides contrasting evidence: it maintains ambiguous valuations and supports the commercialization of self-tracking technologies, while simultaneously preventing the co-optation of the community by commercial values (Barta and Neff, 2016). Significantly, whereas QS pursues individual and community learning, MyData’s means for social change are dependent on success in shaping an ecosystem of new, also commercial, services (see Lehtiniemi and Ruckenstein, 2019). Instigating social change by means of a gradually expanding technical and commercial ecosystem necessitates, for example, demonstrating the benefits for start-ups that aim to occupy niches in it. Commercial values are thus inherent to the sought-after social change. The ideas of individual data agency, their implementation in data technologies and

imagined business benefits come neatly together in the market symmetry frame. The rights-based framing of participation does not bring together commercial interests with an understanding of data agency in equally concrete terms. Commercial data use, then, seems to favor a specific understanding of data and participation: data as an asset for individuals, and data agency as participation in data markets. This understanding, however, leaves potentially narrow parameters for what is at stake; it risks seeing data's value in terms of the competitive dynamics of data markets, and relies on the market to resolve further injustices once the distributive injustice is resolved.

## Conclusion

Data activism only exists in relation to the political economy of personal data and its sociotechnical arrangements. This suggests that commercial potential and alignment with existing interests toward data can have a powerful role in determining the success of data activism's innovations. Our analysis shows how commercial interests involved in data activism can be served by a market framing for data agency and societal participation. The conflation of data agency with the ability to make choices on sharing data can serve firms, but such an approach obviously glosses over the multitude of factors that influence and limit independent choice (Lehtiniemi and Ruckenstein, 2019), and could in the end lead to people sharing more, and more nuanced, personal data. This suggests the course of remaining skeptical of the potential that data activism collaborating with commercial actors has to enhance people's participation in the information society in a sufficient and sustainable manner.

However, it can be difficult for us, as a society, to identify and start resolving data economy's injustices without people's awareness of modes of data collection, access to data, and ability to express choice. While these capabilities are not sufficient for equal participation in the information society, our analysis indicates that they can act as starting points for resolving a variety of economic, sociocultural, and political injustices, provided that data agency is not understood only in terms of data markets and private benefits. This suggests that data activism involving commercial interests can aid in the development of data arrangements that are more just in a sense that surpasses participation in markets, but this may hinge on developing a normative agenda for what participation in a datafied society should involve, and also on articulating nonmarket data agency in concrete terms. Leaving this as an open question may hopefully provide further motivation for scholars to investigate data activism initiatives.


## Acknowledgements


The authors wish to thank Kai Kuikkaniemi and Jogi Poikola for the opportunity to collect data in the MyData 2016 conference, Minna Ruckenstein and other colleagues for their helpful comments at various stages of this research, and the anonymous reviewers for their thoughtful feedback.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research for this article was supported by Tekes, Kone Foundation and Helsingin Sanomat Foundation.

## ORCID iDs

Tuukka Lehtiniemi  <https://orcid.org/0000-0002-9737-3414>

Jesse Haapoja  <https://orcid.org/0000-0001-6877-7957>

## Notes

1. Based on a survey, 40% of conference participants represented firms, 35% the public sector, and the rest NGOs or research institutions. Of firms, half were large corporations and the rest start-ups or small to medium-sized enterprises (SMEs). The representativeness of this sample is, however, questionable.
2. Keynote talks, excluding follow-up discussions, are available at [https://www.youtube.com/playlist?list=PL6\\_IssKYHuPReO0Sr7\\_7GRbUtRkRqnm6m](https://www.youtube.com/playlist?list=PL6_IssKYHuPReO0Sr7_7GRbUtRkRqnm6m)
3. PIMS, personal information management systems, was one of the several names for MyData services.
4. We adopt this abbreviation for Google, Apple, Facebook, and Amazon.

## References

- Andrejevic M (2014) The big data divide. *International Journal of Communication* 8: 1673–1689.
- Baack S (2015) Datafication and empowerment: how the open data movement re-articulates notions of democracy, participation and journalism. *Big Data & Society* 2(2): 1–11.
- Barta K and Neff G (2016) Technologies for sharing: lessons from quantified self about the political economy of platforms. *Information, Communication & Society* 19(4): 518–531.
- Bates J (2013) The domestication of open government data advocacy in the United Kingdom: a neo-Gramscian analysis. *Policy & Internet* 5(1): 118–137.
- Beer D (2018) Envisioning the power of data analytics. *Information, Communication & Society* 21(3): 465–479.
- Benford R and Snow D (2000) Framing processes and social movements: an overview and assessment. *Annual Review of Sociology* 26: 611–639.
- Breindl Y (2013) Assessing success in internet campaigning: the case of digital rights advocacy in the European Union. *Information Communication and Society* 16(9): 1419–1440.
- Brooker K (2018) “I was devastated”: Tim Berners-Lee, the man who created the World Wide Web, has some regrets. *Vanity Fair*, 1 July. Available at: <https://www.vanityfair.com/news/2018/07/the-man-who-created-the-world-wide-web-has-some-regrets>
- Cardullo P and Kitchin R (2018) Being a “citizen” in the smart city: up and down the scaffold of smart citizen participation in Dublin, Ireland. *GeoJournal* 84(1): 1–13.
- Charitsis V, Zwick D and Bradshaw A (2018) Creating worlds that create audiences: theorising personal data markets in the age of communicative capitalism. *tripleC: Communication, Capitalism & Critique* 16(2): 820–834.
- Cinnamon J (2017) Social injustice in surveillance capitalism. *Surveillance & Society* 15(5): 609–625.
- Citron D and Pasquale F (2014) The scored society: due process for automated predictions. *Washington Law Review* 89: 101–133.
- Coll S (2014) Power, knowledge and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society* 17(10): 1250–1263.
- Couldry N and Yu J (2018) Deconstructing datafication’s brave new world. *New Media & Society* 20(12): 4473–4491.



- Crabtree A, Lodge T, Colley J, et al. (2016) Enabling the new economic actor: data protection, the digital economy, and the databox. *Personal and Ubiquitous Computing* 20(6): 947–957.
- Crain M (2018) The limits of transparency: data brokers and commodification. *New Media & Society* 20(1): 88–104.
- Dencik L (2018) Surveillance realism and the politics of imagination: is there no alternative? *Krisis* 1: 31–43.
- Dencik L, Hintz A and Cable C (2016) Towards data justice? The ambivalence of anti-surveillance resistance in political activism. *Big Data & Society* 3(2): 1–12.
- Elder-Vass D (2016) *Profit and Gift in the Digital Economy*. Cambridge: Cambridge University Press.
- Evans B (2017) Power to the people: data citizens in the age of precision medicine. *Vanderbilt Journal of Entertainment and Technology Law* 19(2): 243–265.
- Fligstein N (2001) Social skill and the theory of fields. *Sociological Theory* 29(2): 105–125.
- Fligstein N and McAdam D (2012) *A Theory of Fields*. Oxford: Oxford University Press.
- Fraser N (2008) Abnormal justice. *Critical Inquiry* 34(3): 393–422.
- Garud R (2008) Conferences as venues for the configuration of emerging organizational fields: the case of cochlear implants. *Journal of Management Studies* 45(6): 1061–1088.
- Hess D (2005) Technology- and product-oriented movements: approximating social movement studies and science and technology studies. *Science, Technology & Human Values* 30(4): 515–535.
- Hintz A, Dencik L and Wahl-Jorgensen K (2017) Digital citizenship and surveillance society. *International Journal of Communication* 11: 731–739.
- Iemma R (2016) Towards personal data services: a view on some enabling factors. *International Journal of Electronic Governance* 8(1): 58–73.
- Johnson J (2014) From open data to information justice. *Ethics and Information Technology* 16(4): 263–274.
- Kaun A and Uldam J (2017) Digital activism: after the hype. *New Media & Society* 20(6): 2099–2106.
- Kennedy H (2018) Living with data: aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis* 1: 18–30.
- Lampel J and Meyer A (2008) Field-configuring events as structuring mechanisms: how conferences, ceremonies and trade shows constitute new technologies, industries and markets. *Journal of Management Studies* 45(6): 1026–1035.
- Lehtiniemi T (2017) Personal data spaces: an intervention in surveillance capitalism? *Surveillance & Society* 15(5): 626–639.
- Lehtiniemi T and Ruckenstein M (2019) The social imaginaries of data activism. *Big Data & Society* 6(1): 1–12.
- Lanier J (2013) *Who Owns the Future?* London: Penguin.
- Meyer J and Jepperson R (2000) The “actors” of modern society: the cultural construction of social agency. *Sociological Theory* 18(1): 100–120.
- Milan S and Gutierrez M (2018) Technopolitics in the age of big data. In: Caballero F and Gravante T (eds) *Networks, Movements & Technopolitics in Latin America: Critical Analysis and Current Challenges*. Cham: Palgrave Macmillan, pp. 95–109.
- Milan S and Van der Velden L (2016) The alternative epistemologies of data activism. *Digital Culture & Society* 2(2): 57–74.
- Nafus D and Sherman J (2014) This one does not go up to 11: the quantified self movement as an alternative big data practice. *International Journal of Communication* 8: 1784–1794.
- Nelimarkka M, Kuikkaniemi K, Salovaara A, et al. (2016) Live participation: augmenting events with audience-performer interaction systems. In: *Proceedings of the 2016 ACM conference*

- on designing interactive systems*, Brisbane, QLD, Australia, 4–8 June 2016, pp. 509–520. New York: ACM.
- Nelson J (1993) The study of choice or the study of provisioning? Gender and the definition of economics. In: Ferber M and Nelson J (eds) *Beyond Economic Man*. Chicago, IL: University of Chicago Press, pp. 23–26.
- Pasquale F (2017) Two narratives of platform capitalism. *Yale Law & Policy Review* 35(1): 309–319.
- Poikola A, Kuikkaniemi K and Honko H (2015) *Mydata—A Nordic Model for Human-Centered Personal Data Management and Processing*. Helsinki: Ministry of Transport and Communications.
- Postigo H (2012) Cultural production and the digital rights movement: framing the right to participate in culture. *Information, Communication & Society* 15(8): 1165–1185.
- Pybus J, Coté M and Blanke T (2015) Hacking the social life of big data. *Big Data & Society* 2(2): 1–10.
- Sadowski J (2019) When data is capital: datafication, accumulation and extraction. *Big Data & Society* 6(1): 1–12.
- Schrock A (2016) Civic hacking as data activism and advocacy: a history from publicity to open government data. *New Media & Society* 18(4): 581–599.
- Sharon T and Zandbergen D (2017) From data fetishism to quantifying selves: self-tracking practices and the other values of data. *New Media & Society* 19(11): 1695–1709.
- Silverman D (2006) *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*. London: SAGE.
- Snow D and Benford R (1988) Ideology, frame resonance and participant mobilization. *International Social Movement Research* 1(1): 197–217.
- Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4(2): 1–14.
- Tufekci Z (2014) Engineering the public: big data, surveillance and computational politics. *First Monday* 19. Available at: <https://firstmonday.org/article/view/4901/4097>
- Van Dijck J (2014) Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- West S (2019) Data capitalism: redefining the logics of surveillance and privacy. *Business & Society* 58(1): 20–41.
- Zuboff S (2015) Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30: 75–89.

### Author biographies

Tuukka Lehtiniemi is a PhD candidate in economic sociology. His research interests include data economy and data activism. He works at the Centre for Consumer Society Research at University of Helsinki. He is also affiliated with the Digital Content Communities research group at Aalto University.

Jesse Haapoja is a PhD student at the University of Helsinki's Department of Social Research and works in Aalto University. His background is in social psychology and his interests lie in the intersection of social psychology and information technology.

**Lehtiniemi, T. & Ruckenstein, M. (2019)**  
**The social imaginaries of data activism**  
*Big Data & Society*, 6(1), 1–12



# The social imaginaries of data activism

Tuukka Lehtiniemi<sup>1</sup> and Minna Ruckenstein<sup>2</sup>

Big Data & Society  
January–June 2019: 1–12  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2053951718821146  
journals.sagepub.com/home/bds



## Abstract

Data activism, promoting new forms of civic and political engagement, has emerged as a response to problematic aspects of datafication that include tensions between data openness and data ownership, and asymmetries in terms of data usage and distribution. In this article, we discuss MyData, a data activism initiative originating in Finland, which aims to shape a more sustainable citizen-centric data economy by means of increasing individuals' control of their personal data. Using data gathered during long-term participant-observation in collaborative projects with data activists, we explore the internal tensions of data activism by first outlining two different social imaginaries – technological and socio-critical – within MyData, and then merging them to open practical and analytical space for engaging with the socio-technical futures currently in the making. While the technological imaginary favours data infrastructures as corrective measures, the socio-critical imaginary questions the effectiveness of technological correction. Unpacking them clarifies the kinds of political and social alternatives that different social imaginaries ascribe to the notions underlying data activism, and highlights the need to consider the social structures in play. The more far-reaching goal of our exercise is to provide practical and analytical resources for critical engagement in the context of data activism. By merging technological and socio-critical imaginaries in the work of reimagining governing structures and knowledge practices alongside infrastructural arrangements, scholars can depart from the most obvious forms of critique, influence data activism practice, and formulate data ethics and data futures.

## Keywords

Datafication, social imaginary, data activism, MyData, data ethics, socio-technical futures

This article is a part of special theme on Health Data Ecosystem. To see a full list of all articles in this special theme, please click here: [https://journals.sagepub.com/page/bds/collections/health\\_data\\_ecosystem](https://journals.sagepub.com/page/bds/collections/health_data_ecosystem).

## Introduction

It will not be enough, however, to gain control over the infrastructure of our communicative lives. [...] This is the critical challenge posed by the Big Data era and the new forms of control it ushers in: not simply to reimagine infrastructural arrangements, but also the knowledge practices with which they are associated. (Andrejevic, 2013: 165)

An expanding area of scholarly interest that could be loosely characterized as 'data activism research' explores the harnessing of the capacities of data technology to promote social justice, new forms of agency

and political participation, meanwhile challenging accepted norms, practices and ideological projects (Baack, 2015; Delfanti and Iaconesi, 2016; Greenfield, 2016; Kennedy, 2018; Milan and Gutierrez, 2018; Milan and van der Velden, 2016; Pybus et al., 2015). Data activism research is closely linked with processes

<sup>1</sup>Centre for Consumer Society Research, University of Helsinki, Finland; Department of Computer Science, Aalto University, Finland

<sup>2</sup>Centre for Consumer Society Research, University of Helsinki, Finland

### Corresponding author:

Tuukka Lehtiniemi, Centre for Consumer Society Research, University of Helsinki, P.O. Box 24, 00014 Helsinki, Finland.  
Email: [tuukka.lehtiniemi@iki.fi](mailto:tuukka.lehtiniemi@iki.fi)



of datafication (Ruckenstein and Schüll, 2017; van Dijck, 2014) and the ways in which personal data – any data related to, or resulting from actions by, a person – is being utilized for economic and political aims in an increasingly systematic manner (van Dijck and Poell, 2016; Zuboff, 2015). While data activism research calls for attention to the exploitative forces inherent in processes of datafication, it does not merely detail problematic aspects of datafication; rather, it investigates and draws inspiration from new forms of civic and political engagement that respond to datafication, with the aim of instigating and strengthening more responsible data futures (Milan and van der Velden, 2016).

We build on research that explores how data activism develops ‘alternative social imaginaries’ and creates ‘a new sense for the legitimacy of collective knowledge creation’ (Baack, 2015: 8). The notion of the social imaginary, offered by Taylor (2002), aids in the exploration of how data activists make sense of society’s practices, imagine their social existence, and deal with ‘the expectations that are normally met, and the deeper normative notions and images that underlie these expectations’ (p.106). Developing this idea, Jasanoff (2015) highlights the ‘instrumental and transformative’ role that technology developments play in generating imaginaries of social order, defining *socio-technical* imaginaries as collectively held notions of desirable futures, animated by shared understandings of social aims, and attainable through advances in technology. In terms of data activism, this provides a way to account for the interplay between the design of technologies and the social arrangements that inspire and sustain their production – in other words, how technology both embeds, and is embedded in, the social (Jasanoff, 2015: 2–3). Of particular interest here is how transformations in wider social imaginaries may occur through the development of new practices and associated imaginaries in groups or collectives (Taylor, 2002: 111).

While data activism retains and develops social imaginaries that promote new practices by employing data technology to fulfil aims of social justice or political participation, these capacities can also support opposing perspectives and values. For example, open government data can support liberal democratic values by providing mechanisms for more just governance, but also libertarian agendas by providing justification for privatization and deregulation (Schrock, 2016). It is thus crucial to acknowledge that multiple and conflicting social imaginaries are at work in terms of data activism.

In the following, we discuss tensions arising from alternative ways of ‘framing, packaging, and presenting data’ that ‘have the potential to alter not only our

vision of the world, but also our own theory of knowledge’ (Milan and van der Velden, 2016: 63). Our approach is inspired by Jasanoff’s argument that even if imaginaries are collectively held, ‘multiple imaginaries can coexist within a society in tension or in a productive dialectical relationship’ (2015: 4). We begin by identifying opposing social imaginaries in the context of a single data activist initiative, MyData, and then rework them into a shared dialogue. Our contribution is informed by our involvement in four years of research projects and participatory activities with data activists; it draws from a range of disciplinary sources including critical data studies, anthropology, economic sociology and science and technology studies, and also develops our previous work in the field (Janasik-Honkela and Ruckenstein 2016; Lehtiniemi, 2017; Ruckenstein and Pantzar, 2015).

As a form of data activism, the MyData initiative aims at a more sustainable, and simultaneously citizen-centric, digital economy; it is built on the understanding that people, companies, the public sector and society at large benefit when individuals become more active data citizens and consumers by controlling the gathering, sharing and analysis of personal data. MyData is politically and ideologically thought-provoking by virtue of its self-portrayal as an initiative driven by digital rights, an auto-designation which it introduces as a placeholder in an ambitious aim to provide society with ‘parallel development of digital rights, innovation and business growth’ (Poikola et al., 2015: 4). This translates into the concurrent advancement of processes and policies for protecting individuals’ rights while accommodating the industry’s demands to process personal data in the development of innovative services. MyData seeks to achieve systemic outcomes by rearranging the infrastructure underlying individual-level data practices. The new infrastructure being developed, here understood as technical forms that facilitate user-controlled exchange of personal data (Larkin, 2013), comprises of personal data storages, data schemas and standards, exchange protocols, digital identity frameworks, and permission management tools. The principle of individual data control is intended to be general and sector-independent; indeed, it can be embedded in field-specific initiatives ranging from health and mobility to retail and finances.

In what follows, we separate two social imaginaries: a *technological imaginary* that favours data infrastructure as a corrective measure, and a *socio-critical imaginary* that questions the effectiveness of technological correction. This exercise clarifies political and social alternatives ascribed by different social imaginaries to the data activist initiative, highlighting the need to consider the social structures in play. The more far-

reaching goal, however, is to reach beyond identifying tensions in imaginaries: our account of the unpacking of social imaginaries aims to offer a productive way forward. Towards this end, we finish by discussing how to merge technological and socio-critical imaginaries in the work of reimagining governing structures and knowledge practices alongside infrastructural arrangements.

## Alternative social imaginaries

The perceived need to separate alternative imaginaries before bridging the gap between them was initially triggered by our personal involvement in the MyData initiative. Between 2014 and 2018 we have worked together with developers and data activists in three research projects in fields of health and knowledge work, focusing on personal data uses and emergent data infrastructures. We started our research collaboration with the goal of exploring the wide range of agencies and aims in play in terms of datafication. Initially, we intended to introduce collective aims and expectations in order to open a reflexive conversation about the political and ideological underpinnings of MyData, meanwhile offering ideas on how to promote what we considered societally 'more robust' data activism (see Kennedy, 2018). While our collaboration with data activists was motivated by a mutual understanding that both technology developers and social scientists have an important role in shaping data activism, the first joint meetings were characterized by a certain discomfort. We were viewing a stream of diagrams on PowerPoint slides depicting databases and data flows, in terms of which a more socio-critical imaginary remained oddly irrelevant. Witnessing how society is imagined as being built with information systems shaped our involvement with MyData, pushing us towards an outsider's position from where we had to work our way to a shared dialogue.

We learned first-hand that the socio-critical imaginary that we had internalized through our training in the social sciences, which we also associated with data activism, differed from the technology developers' view, sometimes in a profound sense. In order to explain our position to technology developers we had to clearly spell it out. The socio-critical imaginary is informed by the critical stance characterizing social scientific inquiry, which also questions the optimistic and future-oriented imaginary of technological advances. Drawing from critical political economy and neo-Foucauldian analyses, researchers have explored the effects of datafication on the economy, politics, social life and self-understanding, with particular attention to how technical innovation is outrunning both public understanding and regulation (Kennedy, 2018; Zuboff, 2015). Research highlights

how the introduction of technologies as corrective measures to address identified societal problems leads to new issues that, in turn, need to be corrected: for example, the data economy practices that initiatives like MyData are currently trying to fix were originally justified with a jubilant discourse of the political and societal benefits of online services (West, 2019).

In contrast, the technological imaginary that we encountered in data activism is fed by practical and future-oriented aims. As Fred Turner points out (Logic Magazine, 2017), the engineering attitude includes a tendency to do politics primarily by changing infrastructure. This mindset typically rests on a technolibertarian ideology promoting notions of a free market and autonomous, free-spirited individuals benefiting from advances in information technology (Barbrook and Cameron, 1996; Turner 2006). It tends to take the stance that technology evolution is inevitable: since we cannot stop it, we must make technologies serve us. In this view, information technology, per se, does not generate the undesirable uses to which it is put; rather, they arise when technologies are harnessed to serve particular interests. For technology developers, then, the commitment to societal transformations encompasses ideas of both a more just society, and the correct role of technical means in achieving that transformation (Kelty, 2008). Applying this formulation, initiatives such as MyData treat infrastructural interventions as corrective measures for unsatisfactory societal developments that need to be reversed, or redirected towards fairer and more responsible practices by building new technology. Thus, where the engineering attitude favours infrastructural development, critical scholars, committed to a more socio-critical stance, question the reimagining of such arrangements, particularly if it fails to involve knowledge practices (Andrejevic, 2013).

Our aim is not to paint a caricature of either technology developers or social scientists by separating these social imaginaries, or to claim that either would sufficiently represent any form of data activism. In practice, the two imaginaries are not neatly separable; individual data activists move between them when they explain their future aims. Rather, the two are fundamentally aligned: both recognize the far-reaching consequences of datafication, which enables new approaches for making sense of the world that in turn affect the production of knowledge, business practices and governance (Kitchin, 2014a). In fact, this alignment is what initiated our research collaborations in the first place, as we wanted to better our understanding of the critical potential of technology developers to rework processes related to datafication. The analytical separation between the two can, however, clarify how data activism contributes to 'alternative narratives of

our datafied social reality' (Milan and van der Velden, 2016: 69), and aid in formulation of data activism in terms of social and economic justice (Dencik, 2018). By exposing tensions between social imaginaries, we highlight the contested social aims and expectations of data activism, thereby assisting the evaluation of potential data futures. We argue that the imaginaries inform engagements with new forms of information and knowledge, and their production: that is, they represent dissimilar data futures. In particular, as we suggest below, the imaginaries promoting dissimilar data futures have different relations to the project of individual control of personal data.

### *Participant-observers of MyData*

Studying an initiative such as MyData means dealing with a work-in-progress and uncertain futures in the making. In terms of the actual research process, the emergent nature of the phenomenon at hand has meant that our research has been ethnographically oriented in that we have engaged in ongoing observation and dialogue when interacting with data activists. The observations alerted us to the fact that, rather than being confronted with a uniform 'data activist public', what we face are alternative data futures. To get a better sense of desirable data futures, we started to explore data activists' social imaginaries. This required us to understand how data activists differ from one another, and the nature of their concerns and aims. In the process, however, data activists pushed us not only to study them, but also to offer 'our solution' to remedying the current ills of the data economy. Requests for constructive response echo demands for design input from social scientists in the fields of human-computer interaction and systems design (e.g. Anderson, 1994; Hughes et al., 1994). For us, this meant we needed to 'come up with ingenious solutions to the problem of how to become interesting enough' for data activists and find ways for 'exploring common futures with practices' (Jensen and Lauritsen, 2005: 72–73). Towards this end, we actively had to supply constructive feedback to maintain a productive conversation. In the process, we gained a role in shaping and mobilizing data activism. This has meant that, alongside our research, we have participated in attempts to steer MyData-related improvements constructively. Overall, our engagement with data activism has two aims: to influence data activism by means of our socio-critical imaginary, and to produce scholarly insights providing resources for re-articulating the aims and futures of such activism. Together, these lead to an attempt to synthesize data future visions in a manner that takes the criticism of datafication in the direction that Latour (2004: 247–248) advocates:

the critic should not be 'the one who debunks, but the one who assembles'.

Our empirical material stems from our long-term participation, but also referenced documents, formal and informal interviews, discussions at project meetings, and countless everyday interactions. Alongside the research project's activities, we have taken on participant-observer roles in a 450-member Facebook discussion group<sup>1</sup> consisting of civil servants, activists, technology developers and start-up entrepreneurs. The first author has also participated in the Finnish MyData industry alliance, where a national MyData model is being developed through pilot projects. Further, we have done fieldwork in our roles as organizers, presenters and observers at three annual international MyData conferences in Helsinki since 2016. These collaborations have placed us in the unique position of becoming part of assembling MyData into a socially more robust form of data activism. In the following, we first detail the technological imaginary of MyData activism, the activists' common understanding of relevant issues, and legitimate technological solutions, before moving into more socio-critical understandings of the initiative.

### **'Human-centric' personal data activism**

The high-level MyData vision – described in a white paper<sup>2</sup> written primarily by researchers at the Helsinki Institute for Information Technology and the Tampere University of Technology (Poikola et al., 2015) – outlines a transformation of the 'organization-centric system' into a 'human-centric system' that treats personal data as a resource that the individual can access and control, benefit and learn from. Overall, the MyData vision and related documents suggest that, in the current situation, the collection and analysis of data are too heavily dictated by organizations. As a result, data may be diverted to unforeseen purposes, be combined and analysed in ways that cause people harm or, in another form of loss stressed by MyData developers, may not be used when beneficial to individuals due to the interpretation monopolies of the data-collecting organizations.

The concept of MyData originated with an Open Knowledge Finland working group, where it was developed collaboratively. Open data activists argue that data produced by public authorities should be technically and legally free to use, distribute and reuse (Kitchin, 2014b). According to the MyData initiative, the right to decide on the uses of personal data collected by organizations – such as data on economic transactions, location, smart home appliances, occupational health check-ups or social media – should reside with the data subjects themselves, instead of being monopolized by the organizations. The MyData vision, then,



represents a transformation of the Open Data idea: both aim to release data from a proprietary, monopolistic regime for new uses, but in the case of MyData, both the scope of data, and the scope of benefits derived from data, are scaled down from the collective to the individual level. Jointly formulated MyData principles range from getting access to personal data held by organizations in a machine-readable format – recently also supported by the data portability rights provided in Article 12 of the EU General Data Protection Regulation – to using the data freely, sharing them with third parties, or deleting them. Personal data *become* MyData if they adhere to the spelled-out principles.

Despite the heated debate on ‘who owns personal data’, in the legal sense a logical answer to the question is ‘no one’ (Determann, 2018). In order to avoid the legal debate on data ownership and property rights, MyData activists consciously employ the concepts of data management and control, focusing on individuals’ practical capacity to make use of their data. Figure 1 illustrates how MyData developers perceive their vision. They portray the individual as the ‘operation centre’, placed in the middle of the digital service ecosystem uniting data sources and data endpoints; flows of data pass (either permission-wise or in actual transfers) through the central point.

By aiming to make individuals ‘empowered actors, not passive targets, in the management of their lives both online and offline’ (Poikola et al., 2015: 2), MyData attempts to push the market, or the public sector, to design new services and operation models that allow citizens and consumers to gain personal

value from their data. With the datasets thus created, MyData proponents argue, it is possible to create systems based on real-time feedback, allowing people and organizations to learn about themselves, or readjust their operations. The white paper offers examples of how individuals could utilize personal data for their own purposes, either directly or through sharing. Similar to many other data-driven initiatives, then, MyData promotes new forms of data gathering, sharing and analysis in order to enhance or challenge current practices. Services of this type already exist: for instance, self-tracking devices generate data that people can access. Typically, however, they are problematic as they also utilize personal data for purposes of which their users might not be aware, or knowingly endorse (Crawford et al., 2015; Ruckenstein and Schüll, 2017).

To advance towards its vision, MyData does not affix itself to a particular technological implementation, allowing considerable interpretive flexibility and thereby supporting incommensurable social imaginaries. Indeed, interactions around MyData are characterized by a shared understanding of much-needed technological intervention and, simultaneously, of the complex nature of the issues related to it. MyData is first and foremost an infrastructure-level intervention, focusing on the underlying technological systems needed to realize a ‘human-centric’ personal data ecosystem. Yet, the way it is discussed and promoted has attracted attention in other quarters,<sup>3</sup> from service-developers and tool-makers to policy advocates. Participants are interested in the kinds of information architecture, data exchange standards and

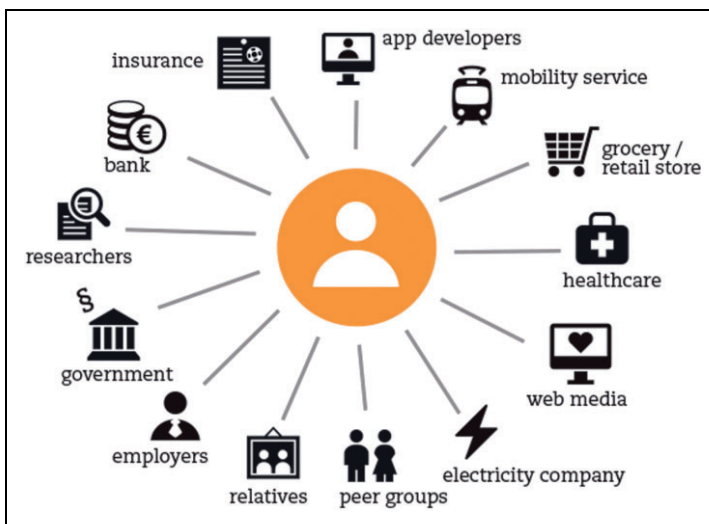


Figure 1. The MyData vision (Poikola et al., 2015).

organizational models needed to support MyData principles, but also the conceptual tools, research and policy required. This tends to attract individuals, companies, or other organizations interested in redefining and readjusting the current data economy by developing approaches giving users more control over their data, including startup companies like Meeco or Cozy Cloud (see Lehtiniemi, 2017), decentralized digital identity technologies such as Sovrin, Kantara Initiative's User-Managed Access protocol, or 'Vendor Relationship Management' systems (see Belli et al., 2017). The approach has some influential supporters in the public sector, as policy makers in Finland<sup>4</sup> and in the European Commission<sup>5</sup> have recognized the potential of the 'human-centric' data management vision. In the next section, we explicate analysis of the technological imaginary underlying MyData, and then discuss the initiative in light of the socio-critical imaginary.

### Reversing the reverse adaptation

By advancing individual empowerment through the control of data collection and data sharing, the MyData vision relies on the ethical principle of 'human self-determination', treating the individual as an autonomous subject with inalienable rights and liberties. The concept of human autonomy, deeply rooted in modern philosophical thinking and embedded in this ethical principle, provided us with one of the first entry points to the ideological underpinnings of the MyData approach (Janasik-Honkela and Ruckenstein, 2016). In essence, MyData can be treated as a practical version of an established philosophical tradition, providing a tool to assess and observe the exploitation of data subjects by a 'system' or 'organizations'. As Taylor (1989) suggests, our perceptions of autonomy and dependency are defined by the notion of free will, according to which an independent agent autonomously sets goals for action. A dependent agent, on the other hand, is someone whose actions are influenced by an external force detached from the individual.

A classic text that resonates with the notion of lost autonomy inhering in the MyData initiative is the treatment of autonomous technology by Winner (1978). Winner perceives the human-technology relationship in terms of Kantian autonomy: via analysis of interrelations of independence and dependence. The core ideas of the MyData vision have particular resonance with Winner's formulation of 'reverse adaptation', wherein the human adapts to the power of the system and not vice versa. Winner presents five methods of action that contribute to reverse adaptation:

- Firstly, the autonomous system, consisting of 'socio-technical aggregates with human beings fully

present, acting and thinking' (Winner, 1978: 242), can take over markets relevant for its operations. According to Winner, markets rarely control the operations of technological systems.

- The second feature of reverse adaptation is that the system strongly influences the political processes that ostensibly regulate its outputs and the prerequisites for its operation. The regulation of markets is so general and non-specific that in reality it is ineffectual.
- The third possible manifestation of reverse adaptation entails the system's finding a 'mission' that fits its technological capabilities. For instance, innovation politics is employed to recognize new objectives or areas of operation to support the market.
- Fourthly, the system might propagate and/or manipulate the needs it serves. As Winner puts it, why wait for public opinion to be shaped when there are numerous ways to influence the formation of social needs?
- Finally, the system might 'run into' a crisis to justify the need for its growth or change; typically, this might be a recognized threat or an alleged deficiency.

Read through a Winnerian lens, MyData is concerned with a gradual loss of control over technological arrangements. Individuals do not have the power to control the system through markets, and the regulatory controls provided, for instance, by data protection and antitrust are insufficient. How MyData frames the problem thus aligns with Winner's classic text, even if the technology developers may not be familiar with the author. The literature that builds the imaginary of technology developers in a more overt manner is typically polemical rather than academic, with close ties to technology circles. For example, Jaron Lanier, a Silicon Valley entrepreneur and pioneer, asks in *Who Owns the Future?* (2013) how to remain human in a society wherein machines appear to be independent agents functioning separately from us:

Popular digital designs do not treat people as being 'special enough'. People are treated as small elements in a bigger information machine, when in fact people are the *only* sources or destinations of information, or indeed of any meaning to the machine at all. (2013: 4, original emphasis)

Lanier associates the current data economy with exploitation and loss of human dignity, as data-gathering entities he calls 'siren servers' control us. He offers monetization of personal data as a solution: 'In a world of digital dignity, each individual will be the commercial owner of any data that can be measured from that person's state or behavior' (2013: 16). In other

words, Lanier promotes commercial symmetry between users and siren servers (p.236) to compensate for the loss of ‘digital dignity’. According to this logic, when commercial agents profit from digital traces, a portion should be distributed to the data subjects as ‘instant remunerations’ in return for data use. A later iteration of the idea refers to ‘data labor unions’ (Arrieta Ibarra et al., 2018) through which users collectively bargain with the data giants.

Like Lanier, the MyData white paper suggests monetization of data as one of the model’s potential benefits (Poikola et al., 2015: 3–4), and we have witnessed discussion of numerous business ideas based on that principle. In this imaginary, more efficient and better-targeted distribution of data generates personal and social advantages by way of economic transactions. Supporting the expansion of the personal data market links MyData principles to value-generating models. Thus MyData is not seen as settling into the existing technology market, but as giving rise to new business models, with economically more balanced use of personal data as their driving force.

When Winner’s ‘autonomous system’, Lanier’s ‘information machine’, or MyData’s ‘organization’ treats humans as mere means to an end, humans are instrumentalized as sources of information instead of being treated as ends in themselves, and what ultimately comes under threat is human dignity. Where Lanier suggests remuneration for personal data as a practical solution for tackling Winnerian reverse adaptation, the promoters of MyData aim at protecting human dignity through advocating MyData principles. Both approaches suggest that people need digital dignity to be capable of self-determination, and argue that dignity can be protected with correctly positioned technology.

### Socio-critical engagement with individual control

From the socio-critical stance, the articulation of citizen and consumer agency in terms of individual-centric data infrastructure is deeply problematic, raising the question of whether MyData actually leads, as its advocates hope, away from reverse adaptation into a more human-centric direction. Or does it, through expanding datafication, encouraging further reliance on data utilization, and opening data to monetization and competition, actually end up *strengthening* the system? Socio-critical engagement with MyData forces us to ask whether it is simply another iteration of Winner’s reverse adaptation. While MyData proposes new data practices based on individual control, it remains ambiguous in how it treats information and knowledge flows. Perceived too simplistically, MyData’s corrective

measures could become a force co-opting rather than countering control of individuals, as with privacy and its protection (Coll, 2014).

Even if MyData activists promote monetization of personal data as only one of many possible technical solutions, the proposition is symptomatic of a belief that individuals can control the market. Promoting the personal data market assumes that people are competent to make informed choices concerning their data (Lehtiniemi and Kortnesniemi, 2017), and that economic rights to data are straightforward determinants of market agency. The notion of a personal data market appeals to the technologically oriented data activists due to the rationale that, since data brokers can successfully monetize personal data for their economic benefit, an intermediary technology could also open the data market to individuals (Belli et al., 2017; Lehtiniemi, 2017). Here, an obvious risk of reverse adaptation lies in the belief that markets ostensibly harnessed to serve individuals would control the system. In other words, a critical imaginary orients us to treat the expanding commodification of personal data as a precarious effort to protect human dignity, but failing to take unpredictable consequences into account. Monetization could potentially lead to further inequalities and discrimination; for instance, privacy might become a prerogative to which only the wealthy can aspire, while the less financially endowed must either trade their personal data, or become data contributors in exchange for basic services such as internet access, housing or electricity. The dividing line could also run along other societal divisions such as technical capability or financial literacy. If new intermediaries start brokering data on behalf of individuals, unprecedented forms of commodifying everyday life might appear: for instance, diseases might become a source of income through data sale. We might face a new class of people responding to the demand by generating data traces and practices that have a market. Moreover, individually optimal data transactions can be socially or societally harmful. The socio-critical imaginary emphasizes contextual aspects of privacy that go beyond the individual: if we consider privacy as a commons (Regan, 2002), individual decisions can erode that commons and harm everyone collectively.

In the MyData 2016 conference,<sup>6</sup> presenters underlined the individual-centricity of the initiative with inventive terms: ‘the Internet of me’, ‘the person as the platform’, ‘the API of me’, ‘the mecosystem’, or the ‘self wide web’. They shared the foundational idea that individuals are interested in controlling personal data. In this respect, the Quantified Self (QS), which took form in 2008, offers an instructive parallel development. The motors of QS are self-trackers, crafting their personal data stories. Individuals are at the

centre of the movement, yet it is not entirely individual-centric. Personal data charts and visualizations trigger collective narration and critical reflection, offering a common language to which people can relate (Nafus and Sherman, 2014). QS has offered support for enquiry into questions of self-knowledge in relation to data practices and the emerging politics of data. As Nafus and Sherman (p. 1877) argue, 'QS is one of the few places where the question of why data matters is asked in ways that go beyond advertising or controlling the behaviors of others.' Due to its infrastructural rather than human-level aims, MyData lacks this kind of collective data work.

We were particularly ready to see MyData in light of critical technology studies, critique of the data economy, and calls for agency (Kennedy et al., 2015), but, in general, our experience with MyData was that while activists are enthusiastic about new perspectives, to be effective they should involve possibilities for technology development or clearly enunciated policy guidance. Where we tended to see a community that would benefit from a more nuanced understanding of its ideological underpinnings, potentially leading to reconsideration of the ways concrete technology projects are envisioned, community members rather considered themselves as practical enablers of technology development. The divide between the social imaginaries concretizes at the point where developers value rapid action and iterations, and social scientists want to take a step back and lean on their concepts and literary sources, resorting to discursive rather than technical intervention in material practices.

Still, there is no doubt that well-executed MyData principles could aid in promoting collective engagement and public culture: for instance, MyData-based approaches encourage the rethinking of governance in companies, as well as advancing new forms of activism. By means of data activism, personal health can be redefined as a collective and political matter; people suffering from serious illnesses can contribute their health data to enhance medical research, or, alternatively, share information about themselves online for everyone to see. The Italian artist Salvatore Iaconesi set up a website featuring medical data related to his brain tumour, alongside a request for 'cures'. By opening a public space within which to experience his illness, he resisted being reduced to the category of a cancer patient constituted by a set of medical data (Delfanti and Iaconesi, 2016). Such examples demonstrate the possibility of re-appropriating personal data and harnessing technological and communicative powers for constructing collective spaces that can call into question existing social and political imbalances. With these observations and experiences, we began to synthesize a more productive relationship between the two social imaginaries.

## Beyond data solutionism

After the publication of the MyData white paper in 2015, the Finnish MyData promoters were contacted by developers, activists and policy-makers in Europe and beyond. Supported by the appeal of the concept, the first MyData conference was held in Helsinki in August 2016. The event brought together 700 participants, differing in interests and objectives, and in the terms and concepts they employed to talk about MyData and similar initiatives. Presentations from various parts of the world and different sectors of society showcased services and tools that either explicitly follow MyData principles or, without committing to any form of data activism, shared its political aims. According to a key promoter, the conference was an occasion where the 'MyData community started to become self-aware' (MyData.org, 2017: 16). In a summary speech for the conference, Valerie Peugeot (2016) from Orange Labs pushed the audience to widen their imaginary by introducing MyData as a social movement and expanding the activist stance beyond technological and regulatory issues. In light of the imaginaries we have outlined, Peugeot's summary indicated that, when viewed through the technological imaginary, MyData is an ambitious political project advancing human-centricity, but in terms of the socio-critical imaginary it is not ideological enough to reach its aims of digital dignity, empowerment and citizen-centricity. Building on this idea, in order to become sufficiently ideological, MyData should more explicitly outline intended aims for technology development, including desired and undesired objectives of data usage. The infrastructure-level vision should be combined with actual knowledge practices and clearly enunciated outcomes. At the least, it should propose how to move beyond defining personal data within an individual property paradigm, and take into account the relations and politics that uses of personal data bring into being. As argued above, with its ambiguous stance towards information and knowledge production, as long as it conforms to individual data control, MyData can merely introduce new forms of exploitation. In order to avoid this, more clearly stated societal aims are needed.

In August 2017, the authors teamed up with Peugeot and co-hosted a track called 'Our Data' at the MyData conference to promote the reimagining of knowledge practices alongside infrastructural data arrangements. By talking, somewhat provocatively, about 'our' instead of 'my' data, we promoted collective engagement through data activism with the intent of combining technology-oriented MyData activism with a socio-critical stance on the individual-centricity of the initiative. We argued that developing data technologies for the individual and leaving it up to the market to

correct the economic imbalances will hardly work alone (Lehtiniemi, 2017). Technologically savvy data activists are urgently needed for clarifying and mediating the work that data practices and infrastructures require. Therefore, in order for the socio-critical imaginaries to be realized, technological designs and repositionings of the data infrastructure are required to strengthen forms of activism not centred on the individual or on data, but on people collectively as the sources and distributors of data. The presentations and subsequent conversations focusing on 'Our Data' in the MyData conference and beyond suggest several possibilities for combining the technological and socio-critical imaginaries, thereby bringing back together the imaginaries that we initially separated for analytical purposes. However, rather than merely traveling a full circle back to Jasanoff's notion of socio-technical imaginary, the analytic separation and ensuing merging of the social imaginaries has allowed us to open practical and analytical space for the exploration of the socio-technical future currently in the making. From that space we can influence data activism practice and see more clearly what is timely in terms of data activism research and data ethics.

The following outcomes of combining technological and socio-critical imaginaries operate on different scales and registers, and are provided here as examples of future data activist work. Together they indicate growing interest among the MyData participants in working towards consciously building a socio-technical future and thinking beyond data solutionism. First, services abiding by MyData principles could exercise notions of desired and undesired data use by means of collective data governance. The concept of governance is already built into the MyData vision; the developers are cognizant that a functioning digital service environment requires that interoperability is assured by rules that govern both technical and operational aspects of data flows. These rules could be coupled with explicit governance of data usage and exploitation, aligning MyData principles with collectively agreed notions of acceptable data use. At one end, personal data that is acknowledged as a constitutive part of personal identity (Floridi, 2017: 95–96) could be considered strictly off limits in terms of trading or processing. At the other end, it could be agreed that some data may be safely shared with almost anyone. In practice, activist work is needed to explore how to reach decisions collectively about these extremes and the space in between.

A second combination takes advantage of infrastructural technologies' relying on MyData principles in producing data commons, which can be formed of proprietary personal data, but can also bring together other kinds of open data sources benefiting collective aims. In this way, MyData can aid specific

collectives in reclaiming personal data to benefit the community at large instead of the individual, adding a societally oriented layer to technological infrastructure. Various projects working towards the creation of data commons already exist: for instance, for platform cooperatives (Carballa Smichowski, 2016), in the context of the smart city (Morozov and Bria, 2018), in the health research realm (Evans, 2017), and through data sharing platforms such as Open Humans (Ball, 2018).

Third, the tradition of cooperative-based governance can function as a basis for shared data ownership and citizen-led initiatives by promoting digital rights. In Europe, Nordic countries in particular, sharing the goal of advancing a more responsive and responsible digital society can provide a rich cultural breeding ground for MyData principles (Janasik-Honkela and Ruckenstein, 2016), preventing society being subordinated to proprietary and monopolistic data infrastructures. Cooperative ownership models are undergoing experimentation in initiatives such as Healthbank and MIDATA.coop. Overall, MyData technology developers could collaborate with social movements aiming to solve societal problems in order to demonstrate, in practice, how data can shape knowledge practices and generate advocacy and public benefits (Milan and Gutierrez, 2018).

Finally, refusals to share personal data can also become political acts and corrective measures. Despite our focus on proactive (Milan and van der Velden, 2016) data activism that sees the beneficial potential of datafication, there is a need for continued questioning of the naturalization of relentless data gathering and storing, and to insist that less data are gathered for unknown future uses. The very collection of data, and not only subsequent uses of it, may have negative implications. In line with our reasoning, reactive data activism – for instance, refusing certain forms of sharing personal data with corporations or the state (Moore and Robinson, 2016) – can also be technologically mediated, engaged in collectively, and leveraged for the collective good.

## Conclusion

An important consideration for data activism is how to ensure a robust enough conceptual grounding for advancing the public good, as concepts such as political participation, privacy, autonomy, or health are taken up, enacted, and altered through interactions with data processing technologies, and become enmeshed with engineering and design. This calls for rigorous analyses of the production of data infrastructures, how they are imagined, and of what kinds of ideological and everyday data relations they consist. In this article, we have

contributed to this debate by separating two social imaginaries at stake in terms of data activism – the technological and the socio-critical – and highlighting their discrepancies. Then we brought the two together in order to open practical and analytical space for engaging with socio-technical futures and promoting dialogue across professional and scholarly fields.

In becoming a part of the data activist scene that we are studying, we are participants in a widening scholarly trend. In order to understand the aims and tensions of data activism in an empirically grounded manner, researchers have begun to explore applied perspectives, often by collaborating with data activists and data-driven initiatives outside their academic spheres. Applied research perspectives can deepen understandings of datafication by revealing how data technologies are taken up, valued, and repurposed in ways that either do not comply with imposed data regimes, or that mobilize data in alternative or inventive ways. For instance, health-related initiatives have strong ties with the academic community in addressing the tension between data openness and data ownership and asymmetries in terms of data usage and distribution (Kish and Topol, 2015; Nafus, 2016), and the inadequacy of informed consent and existing privacy protections (Sharon, 2016a). A shared aim is also the re-articulation of ethically motivated concepts, such as sharing, solidarity, commons and the public good (Prainsack and Buyx, 2017; Sharon, 2016b).

All along we have argued that the rapidly changing technology landscape calls for taking Latour's (2004) view on critique seriously: we need to keep asking what productive 'critical engagement' means in the context of data activism and developing data infrastructures. Based on our experience with MyData, it seems that in order to succeed in cross-professional dialogue, social scientists need to exercise disciplinary self-awareness; they need to understand how their socio-critical imaginary differs from the imaginary of technology developers and be ready to depart from the most obvious forms of critique associated with the exploitative forces of datafication. By offering critique that is pieced together in a constructive manner, data activism research can focus on collectively sustainable socio-technical data futures. As we have demonstrated, by uncovering the aims and contestations around data activism, socio-critical imaginaries can aid in promoting progressive 'public good agendas', offering support for navigating policy-crafting, technology companies' proprietary software, and data platforms that have become participants in deciding what counts in people's lives. Linking knowledge production to data activism practice, we can strengthen the understanding of how data technologies become part of everyday practices with societally comprehensive goals.

## Acknowledgements

We thank Nina Janasik, Valerie Peugeot, and the members of the MyData community, especially Kai Kuikkaniemi and Antti "Jogi" Poikola, for generously sharing their insights and aims. We also thank the three anonymous reviewers for their constructive comments, and Tamar Sharon and Federica Lucivero for bringing the special issue together.

## Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Tekes – the Finnish funding Agency for Innovation, grant no. 2676/31/2015.

## Notes

1. Critique of the data economy is indeed discussed on Facebook, even though some activists refuse to use the platform.
2. The 'white paper' is a summary of a more comprehensive Finnish-language study commissioned by the Ministry of Transport and Communication (Poikola et al., 2014).
3. See <http://www.mydata.org> (accessed 13 December 2018).
4. In 2015, the Finnish government programme included the following aim: 'People's right to decide about and monitor their personal information will be enhanced, while ensuring the smooth transfer of data between the authorities' (see Prime Minister's Office, 2015: 27).
5. The European Commission organized a roundtable for personal information management service developers: <https://ec.europa.eu/digital-single-market/en/news/emerging-offer-personal-information-management-services-current-state-service-offers-and> (accessed 13 December 2018).
6. <http://www.mydata2016.org> (accessed 13 December 2018).

## References

- Anderson R (1994) Representations and requirements: The value of ethnography in system design. *Human-Computer Interaction* 9(3): 151–182.
- Andrejevic M (2013) *Infoglut: How too much information is Changing the Way We Think and Know*. New York, NY / London: Routledge.
- Arrieta Ibarra I, Goff L, Jiménez Hernández D, et al. (2018) Should we treat data as labor? Moving beyond 'free'. *American Economic Association Papers & Proceedings* 108: 38–42.
- Baack S (2015) Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data and Society* 2(2): 1–11.
- Ball M (2018) *From personal data to collective power*. Presentation at MyData 2018 conference. Available at:

- [https://www.youtube.com/watch?v=MOT\\_fj4LP8Q](https://www.youtube.com/watch?v=MOT_fj4LP8Q) (accessed 13 December 2018).
- Barbrook R and Cameron A (1996) The Californian ideology. *Science as Culture* 6(1): 44–72.
- Belli L, Schwartz M and Louzada L (2017) Selling your soul while negotiating the conditions: From notice and consent to data control by design. *Health and Technology* 7(4): 453–467.
- Carballa Smichowski B (2016) Data as a common in the sharing economy: A general policy proposal. *Document de travail du CEPN* 2016-10.
- Coll S (2014) Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. *Information, Communication and Society* 17(10): 1250–1263.
- Crawford K, Lingel J and Karppi T (2015) Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies* 18(4–5): 479–496.
- Delfanti A and Iaconesi S (2016) Open source cancer. Brain scans and the ritual of biodigital data sharing. In: Barney D, Coleman G, Ross C, et al. (eds) *The Participatory Condition in the Digital Age*. Minneapolis: University of Minnesota Press, pp. 123–143.
- Dencik L (2018) Surveillance realism and the politics of imagination: Is there no alternative? *Krisis* 1: 31–43.
- Determann L (2018) No one owns data. *UC Hastings Research Paper* 265: 1–49.
- Evans B (2017) Power to the people: Data citizens in the age of precision medicine. *Vanderbilt Journal of Entertainment and Technology Law* 19(2): 243–265.
- Floridi L (2017) Group privacy: A defence and an interpretation. In: Taylor L, Floridi L and van der Sloot B (eds) *Group Privacy. New Challenges of Data Technologies*. Cham: Springer, pp. 83–100.
- Greenfield L (2016) Deep data: Notes on the n of 1. In: Nafus D (ed.) *Quantified: Biosensing Technologies in Everyday Life*. Cambridge, MA: MIT Press, pp. 121–146.
- Hughes J, King V, Rodden T, et al. (1994) Moving out from the control room: Ethnography in system design. In: *Proceedings of the 1994 ACM conference on computer supported cooperative work*, Chapel Hill, NC, October 22–26. New York, NY: ACM, pp.429–439.
- Janasik-Honkela N and Ruckenstein M (2016) My data: Teknologian orjuudesta digitaaliseen vastarintaan. *Tieteessä tapahtuu* 34(2): 11–19.
- Jasanoff S (2015) Future imperfect: Science, technology and the imaginations of modernity. In: Jasanoff S and Kim S-H (eds) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press, pp. 1–33.
- Jensen C and Lauritsen P (2005) Qualitative research as partial connection: Bypassing the power–knowledge nexus. *Qualitative Research* 5(1): 59–77.
- Kelty C (2008) *Two Bits: The Cultural Significance of Free Software*. Durham, NC: Duke University Press.
- Kennedy H (2018) Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis* 1: 18–30.
- Kennedy H, Poell T and van Dijk J (2015) Data and agency. *Big Data and Society* 2(2): 1–7.
- Kish L and Topol E (2015) Unpatients: Why patients should own their medical data. *Nature Biotechnology* 33(9): 921–924.
- Kitchin R (2014a) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 1–12.
- Kitchin R (2014b) *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. London: Sage.
- Lanier J (2013) *Who owns the Future?* London: Penguin Books.
- Larkin B (2013) The politics and poetics of infrastructure. *Annual Review of Anthropology* 42: 327–343.
- Latour B (2004) Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry* 30(2): 225–248.
- Lehtiniemi T (2017) Personal data spaces: An intervention in surveillance capitalism? *Surveillance and Society* 15(5): 626–639.
- Lehtiniemi T and Kortensniemi Y (2017) Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach. *Big Data and Society* 4(2): 1–11.
- Logic Magazine (2017) Don't be evil: Fred Turner on utopias, frontiers, and programmers. *Logic Magazine* 3. Available at: <https://logicmag.io/03-dont-be-evil/> (accessed 13 December 2018).
- Milan S and Gutierrez M (2018) Technopolitics in the age of Big Data. In: Caballero F and Gravante T (eds) *Networks, Movements & Technopolitics in Latin America: Critical Analysis and Current Challenges*. Cham: Palgrave Macmillan, pp.95–109.
- Milan S and van der Velden L (2016) The alternative epistemologies of data activism. *Digital Culture and Society* 2(2): 57–74.
- Moore P and Robinson A (2016) The quantified self: What counts in the neoliberal workplace. *New Media and Society* 18(11): 2774–2792.
- Morozov E and Bria F (2018) *Rethinking the Smart City. Democratizing Urban Technology*. New York, NY: Rosa Luxemburg Stiftung.
- MyData.org (2017) *MyData 2017 end report*. MyData.org. Available at: [https://issuu.com/mydataorg/docs/end\\_20report\\_20mydata\\_202017\\_20\\_28d](https://issuu.com/mydataorg/docs/end_20report_20mydata_202017_20_28d) (accessed 13 December 2018).
- Nafus D (ed.) (2016) *Quantified: Biosensing Technologies in Everyday Life*. Cambridge, MA: MIT Press.
- Nafus D and Sherman J (2014) This one does not go up to 11: The quantified self movement as an alternative Big Data practice. *International Journal of Communication* 8: 1784–1794.
- Peugeot V (2016) *Summary Talk of the Conference*. Presentation at MyData 2016 conference. Available at: <https://www.youtube.com/watch?v=3rcYOeiHSxk> (accessed 13 December 2018).
- Poikola A, Kuikkaniemi K and Honko H (2015) *MyData – A Nordic Model for Human-centered Personal Data Management and Processing*. Helsinki: Finnish Ministry of Transport and Communications.

- Poikola A, Kuikkaniemi K and Kuittinen O (2014) *My Data – Johdatus ihmiskeskeiseen henkilötiedon hyödyntämiseen*. Helsinki: Finnish Ministry of Transport and Communications.
- Prainsack B and Buyx A (2017) *Solidarity in Biomedicine and Beyond*. Cambridge: Cambridge University Press.
- Prime Minister's Office (2015) *Finland, a Land of Solutions: Strategic Programme of Prime Minister Juha Sipilä's Government*, 29 May. Available at: [http://valtioneuvosto.fi/documents/10184/1427398/Ratkaisujen+Suomi\\_EN\\_YHDISTETTY\\_netti.pdf/8d2e1a66-e24a-4073-8303-ee3127fbfcac](http://valtioneuvosto.fi/documents/10184/1427398/Ratkaisujen+Suomi_EN_YHDISTETTY_netti.pdf/8d2e1a66-e24a-4073-8303-ee3127fbfcac) (accessed 13 December 2018).
- Pybus J, Coté M and Blanke T (2015) Hacking the social life of big data. *Big Data and Society* 2(2): 1–10.
- Regan P (2002) Privacy as a common good in the digital world. *Information, Communication & Society* 5(3): 382–405.
- Ruckenstein M and Pantzar M (2015) Datafied life: Techno-anthropology as a site for exploration and experimentation. *Techné: Research in Philosophy and Technology* 19(2): 191–210.
- Ruckenstein M and Schüll N (2017) The datafication of health. *Annual Review of Anthropology* 46(1): 261–278.
- Schrock A (2016) Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society* 18(4): 581–599.
- Sharon T (2016a) Self-tracking for health and the quantified self: Re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. *Philosophy and Technology* 30(1): 1–29.
- Sharon T (2016b) The Googlization of health research: From disruptive innovation to disruptive ethics. *Personalized Medicine* 13(6): 563–574.
- Taylor C (1989) *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
- Taylor C (2002) Modern social imaginaries. *Public Culture* 14(1): 91–124.
- Turner F (2006) *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago, IL: University of Chicago Press.
- Van Dijck J (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- Van Dijck J and Poell T (2016) We Understanding the promises and premises of online health platforms. *Big Data and Society* 3(1): 1–11.
- West S (2019) Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society* 58(1): 20–41.
- Winner L (1978) *Autonomous Technology – Technics-out-of-Control as a Theme in Political Thought*. Cambridge, MA: MIT Press.
- Zuboff S (2015) Big Other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30: 75–89.







**UNIVERSITY  
OF TURKU**

ISBN 978-951-29-8001-7 (PRINT)  
ISBN 978-951-29-8002-4 (PDF)  
ISSN 0082-6987 (Print)  
ISSN 2343-3191 (Online)