

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

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.
- 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.
- 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.
- 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:10.1162/15424760322493212.
- 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. ‘Hypernudge’: 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>.