Journal of Business Ethics (2021) 168:565–578
https://doi.org/10.1007/s10551-019-04239-z

ORIGINAL PAPER

The General Data Protection Regulation in the Age of Surveillance Capitalism

Jane Andrew[1] · **Max Baker**[1]

Received: 18 March 2018 / Accepted: 12 June 2019 / Published online: 18 June 2019

Abstract
Clicks, comments, transactions, and physical movements are being increasingly recorded and analyzed by Big Data processors who use this information to trace the sentiment and activities of markets and voters. While the benefits of Big Data have received considerable attention, it is the potential social costs of practices associated with Big Data that are of interest to us in this paper. Prior research has investigated the impact of Big Data on individual privacy rights, however, there is also growing recognition of its capacity to be mobilized for surveillance purposes. Our paper delineates the underlying issues of privacy and surveillance and presents them as in tension with one another. We postulate that efforts at controlling Big Data may create a trade-off of risks rather than an overall improvement in data protection. We explore this idea in relation to the principles of the European Union's General Data Protection Regulation (GDPR) as it arguably embodies the new 'gold standard' of cyber-laws. We posit that safeguards advocated by the law, anonymization and pseudonymization, while representing effective counter measures to privacy concerns, also incentivize the use, collection, and trade of behavioral and other forms of de-identified data. We consider the legal status of these ownerless forms of data, arguing that data protection techniques such as anonymization and pseudonymization raise significant concerns over the ownership of behavioral data and its potential use in the large-scale modification of activities and choices made both on and offline.

Keywords General Data Protection Regulations · Privacy · Surveillance · Ethics · Big Data · Technology

Introduction

The proliferation of Big Data analytics used in social media, software logs, and for tracking financial transactions has already posed significant ethical challenges to governments and corporations across the globe. Recently, Big Data has been implicated in two scandals involving Cambridge Analytica where large data sets were used to identify and target voters during both the US presidential election and the UK's Brexit decision (Scott 2018). These high-profile incidences have raised awareness of some of the complex issues surrounding Big Data. The collection, utilization, and trading of Big Data now traverses international jurisdictions, and different legal and ethical frameworks. A case in point is the legal action taken by privacy activist Max Schrems against Facebook Ireland, where it was found that the company violated data protection laws by 'exporting' data to its US-based parent company. This case is one of many representing a European-led legal challenge to the standard practices ('model clauses') used by global tech companies when collecting the personal data of half a billion EU citizens. The regulatory response to Big Data in the EU has also been progressively strengthened, with the European Commission's General Data Protection Regulation (GDPR) emerging to address mounting concerns about the collection and use of Big Data. This new 'gold standard' of data protection law will impose costly changes to the practices of all companies dealing with data related to European citizens, and carries with it the potential of significant fines and penalties for non-compliance.

In the main, prior ethics-based research on Big Data has adopted a jurisprudence reasoning wherein the *right to privacy* has become the "preeminent mobilizing concept" in the field (Lyon 2014, p. 10). However, broader contextual issues in relation to changes in control and power ushered in by Big Data are also beginning to receive attention.

✉ Max Baker
max.baker@sydney.edu.au

[1] Discipline of Accounting, The University of Sydney Business School, H69 - Economics and Business Building, Sydney, NSW 2006, Australia

There is a burgeoning field of work that discusses *surveillance capitalism* or *data capitalism* and the development of information networks between various governments and companies that aim to aggregate, trade, and use data to influence both individuals and groups (Gandy Jr 2012; Lyon 2014; Marwick 2012; West 2017; Zuboff 2015, 2019). This paper attempts to contribute to a growing body of work that explores the junction between the ethical issues of privacy and surveillance (Andrejevic 2014; Andrejevic and Gates 2014; Lyon 2003, 2014; Van Dijck 2014). Specifically, we tease out the underlying assumptions of 'privacy' and 'surveillance' risks and how each conceives of individual 'harm.' In doing so, we consider prior discussions surrounding tensions between similar constructs, such as Etzoni (Etzioni 2018) and Newkirk (2018) who discuss the difficulties in reconciling national security and privacy concerns. These debates point to the importance of clear legal frameworks in minimizing harms in often complex situations which require a clarity around "whose rights are curbed, to what degree, and how does a society decide which side to favor?" (Newkirk 2018, p. 13).

This paper argues that while privacy and surveillance are often interrelated, there are also situations, often created by laws and regulations, where these risks are in tension. Here the attempts to address privacy concerns may lead to a 'trade-off' that introduces greater surveillance risk. We explore this 'trade-off' in relation to the principles and recommendations of the new GDPR, which is likely to influence laws in other jurisdictions such as the US (Safari 2016). Moreover, as a law of general protection rather than of strictly privacy, the GDPR has been designed with the challenges of surveillance in mind (WP 55, 215, 221, 228, 237), thus it is important to understand whether this aim is being achieved. Our focus is on the treatment of data modification techniques (anonymization and pseudonymization) under the law and their ability to intensify surveillance risks. With reference to Zuboff's (2019) latest book, *The Age of Surveillance Capitalism*, we argue that while efforts to de-identify data subjects are no doubt beneficial in terms of privacy, these same processes, which disconnect subjects from their personal data, free up the use and trade of other types of *behavioral data*. It is important to note here that these techniques represent a new form of surveillance that is less to do with a centralized state-led panopticon targeting individuals as private citizens and more concerned with promoting various data marketplaces engaged in the trade of both behavioral data and its associated predictive algorithms and products. Our concern, following Zuboff (2019), is that the freedom of these data products to operate, largely outside the GDPR, has serious ethical consequences for their use in the modification of behavior and choices of individuals as consumers and voters.

The next section defines Big Data and explores it as a potential site of privacy and surveillance risks. We then conceptually delineate privacy and surveillance as two separate but related data protection issues, both of which are always present in various Big Data practices. Our paper then introduces the relevant features and aspects of the GDPR and contrasts its treatment of *identified* data vis-à-vis de-identified data, that is, data that have been anonymized or pseudonymized. We first outline the improvements to privacy and surveillance risks, created by the GDPR, in terms of identified data. However, our main focus is on how de-identification techniques, rather than improving data protection outright, promote practices associated with surveillance risk. In the discussion of the paper, we extend the recent work of Zuboff (2019) and introduce the concept of *res nullius* (objects with no owners) to explore the ethical issues associated with de-identified data sets. Finally, we conclude the paper by outlining the broader ethical concerns of the GDPR and the aspects of the law that may offer some promise for the future of data protection.

What is Big Data?

The term 'Big Data' refers to the binding of advanced predictive tools with large data sets on human activity, with the aim of tracking, monitoring, analyzing, and disseminating this information (Boyd and Crawford 2012; Lohr 2012; Warner and Sloan 2014). Ironically, the defining feature of Big Data is not its size alone but a capacity to search and integrate large data sets in order to more accurately predict future events, activities, and behaviors (Boyd and Crawford 2012). There are three main sources of Big Data. The first is organizational records, which include economic transactions, contracts, or registrations with companies or governmental institutions such as births, deaths, and marriages. The second is accumulated through the use of websites and includes e-commerce activity, exchanges on social media platforms, and the use of browser searches. This type of information can include pictures, videos, and text comments on various apps and sites (Qualman 2010). The third source of Big Data relates to the physical movement of people and is collected through private and public surveillance cameras, smartphones, and satellites. These devices are now often equipped with facial (and other) recognition technology to track individual movements. Indeed, the growing number of sensors embedded in objects, bodies, and places intensifies the potential for surveillance of individuals (Miraz et al. 2015).

The properties that define Big Data, such as volume, variety, and veracity (Geng 2017), contain insights into the advances of these new technologies and practices. The exponential decrease in the cost of processing and memory

has drastically increased the volume of data in circulation, which is now measured in zettabytes. More than the quality or completeness of data, the usefulness of Big Data depends on interconnection with other data sets. In addition to volume, the variety of text, images, and video produced around human activities and their digital footprints is an important feature of Big Data. This includes the production of new types of metadata (data that describe and give information about other data), which has begun to change the shape and usefulness of large data sets. Lastly, the speed by which Big Data is accumulated and processed, known as velocity, has grown considerably. Here, advances in automation have meant that data processing increasingly bypasses the need for human intervention. All of these qualities present new and interesting legal and ethical challenges for those either using Big Data or subject to its practices.

The Risks of Big Data

As Rouvroy (2016, p. 3) has argued, the production and consumption of Big Data produces new legal and ethical risks. Clearly, it creates new forms of individual visibility that make it possible "to model human behavior and predispositions for various purposes" (Rouvroy 2016, p. 3). Often, this is based on data collected from the fairly routine contexts in which people live—our supermarket and online purchases, our google searches, and our physical movements within our cities. In what follows, we review two ways that ethics-based research has engaged with Big Data, that is, through exploring privacy and surveillance issues. Privacy work primarily deals with the impact of Big Data practices on individual rights, including the extraction of data, the use of disclosures, and consent, access, and privileges of users. On the other hand, surveillance research has tended to describe a broader apparatus of power and the changing relationship between organizations, the state, and populations. No doubt both these areas have many intersections. For instance, direct surveillance activities by government agencies often involve some level of privacy invasion. However, for the purposes of our analysis, we attempt to identify the specific sets of issues and questions dealt with by each *concept,* rather than by the streams of research relevant to these concepts. We refer to these concepts as *privacy risk*, which focuses on the adverse effects of Big Data practices on the rights of individuals as private citizens, and *surveillance risk*, which relates to how Big Data is used by the state and companies to monitor and control individuals as participants in politics and markets. This distinction will be further developed in the following parts and will form the basis of a later critique of GDPR.

Privacy Risk

Privacy can be thought of as "a claim, entitlement, or right of an individual to determine what information about himself or herself may be communicated to others" (Torra 2017, p. 5). The original exploration of privacy as a legal right was made by Warren and Brandeis (1890) who defined privacy as a protection of an individual and the right "to be let alone."[1] In the contemporary digital world of Big Data, this right "to be let alone" is more difficult to conceptualize, but in general it reflects the right to privacy surrounding digital information that pertains to personal attributes of individuals. The focus on individual data privacy owes its legacy to American jurisprudence scholarship (Torra 2017), as well as to the development of work in other areas, such as communication, technology, sociology, psychology, and economics (Hull 2015; Kokolakis 2017; Taddicken 2014; Young and Quan-Haase 2013). However, some have argued that research into privacy has prioritized a concern for individual rights in relation to threats from "outside," often distant, powerful interest groups such as the state or companies (Floridi 2014; Taylor et al. 2016). This has created an ontological preference for individual rights vis-à-vis the rights of groups—and is a particular blind spot especially considering the right to organize and the rights of nations to self-determination. Moreover, recent developments in social contract theory, maintain that a proper understanding of privacy requires an acknowledgement of dialogues around not only rights but also *norms* (Martin 2016a). Norms here refer to important contextual factors, outside legal environments, and broader political and social discourses (Martin and Nissenbaum 2016; Nissenbaum 2009). It is norms, in concert with rights, that form the basis of both explicit and implicit social contracts which govern the practice and obligation of sharing information (Martin 2016b).

For the purposes of this paper, from now on, when we talk about privacy we are discussing it within the context of digital personal data such as name, email addresses, phone numbers, credit cards, IP addresses, and information pertaining to personhood such as background, age, and religion. There is little doubt that the use of Big Data can have positive outcomes in social policy (Blumenstock et al. 2015), targeting criminal behavior (Berk 2012), in optimizing medical and other health care delivery (Obermeyer and Emanuel 2016), and in addressing inequality and poverty (Glaeser et al. 2016, 2018)—but at the same time, privacy risks have grown. The PEW Research Center's report on public perceptions of privacy indicates that 91% of US adults believe they have lost control over their personal data (Madden 2014).

[1] The pithy stave "to be let alone" was originally coined by the then US Supreme court justice Thomas Cooley (1888).

These concerns are not unfounded; under US law, companies do not need consent to collect personal data nor do they have to share it with these users on request (The Electronic Communications Privacy Act, 18 U.S.C. §2510). In response to these growing concerns, researchers have examined the harm caused to online users from the lost privacy associated with Big Data (Newman 2013); users' expectations related to privacy (Martin and Shilton 2016); the role of privacy preferences (Baruh and Popescu 2017); the impact of privacy notifications (Martin 2015b, 2016b); and the possibilities of informed consent as a mechanism to change user behavior (Barocas and Nissenbaum 2014).

While practices like 'informed consent' have arisen to empower participants in light of earlier abuse, the legal apparatus to handle ownership of personal photos, videos, and text that are posted to sites remains a complex and developing domain (Madden et al. 2017; Pagallo 2017). Martin (2016a, p. 51, b) argues that the current legal emphasis on disclosures of terms and conditions associated with data use has paradoxically freed firms "to implement questionable privacy practices as long as the practices are accurately reported." In addition, disclosure statements are often unidirectional, giving users little option to change or influence the actual practices of companies, like Facebook, which remains "uniquely positioned to undercut … expectations of privacy" (Martin 2016a, p. 52). However, it is important to note that some regulatory frameworks are elevating the rights of data subjects. For instance, the European GDPR has been the source of growing academic debate as it represents new and significant restrictions for those organizations using European data, including the changes related to remedies, fees, restrictions, enforcement, and organizational practices for establishing consent and trust around the collection and general use personal data (Chaudhuri 2016; Kamara 2017; Koops and Leenes 2014; Rotenberg and Jacobs 2013; Safari 2016; Wachter 2018; Zarsky 2016).

While in technical terms no technology can guarantee perfect privacy online—even for the most sophisticated and cautious practitioners (Calmon et al. 2015)—recent research has explored the efficacy and utilization of techniques like anonymization to alleviate privacy risk. These techniques work by disconnecting or 'de-identifying' a data subject from a data set (Torra 2017). They are no doubt appealing, as they offer regulators and data controllers various methods to avoid violating privacy rights while freeing up the data for future use and analysis (Gritzalis 2004). However, despite the potential of these methods to address privacy risk, there may still be broader ethical issues and risks surrounding the way de-identified data are used to monitor and predict the behaviors of subjects. So, while privacy concerns may be allayed when information ceases to relate to individuals specifically, the effect of Big Data on society remains a significant concern. We will now explore these issues through the concept of data *surveillance risk*, which points to the ways Big Data can be mobilized to change the power relationship between state, companies, digital elites, and governed populations.

Surveillance Risk

While privacy continues to be a significant concern, more recently, the relationship between Big Data and surveillance has stimulated considerable research. This work conceives of Big Data as an 'industry' or social, political, and economic system, rather than simply a set of technological practices (Martin 2015a). Modern data surveillance is commonly thought to have started with the development of internet commerce in the late 1990s, through ushering in a "new scope and scale of tracking" and "data collection practices" (West 2017, p. 6), which were later refined by a few US companies such as Google and Facebook (Englehardt and Narayanan 2016). Surveillance risk is concerned with the accumulation of personal and non-personal information by these large organizations. Here, non-personal information refers to behavioral data known as User Profile Information (UPI) which includes, search queries, clicks, and the time spent on each website. New technologies are now able to infer personal information from UPI or behavioral information, and thus a knowledge of individual activities and tendencies can be accumulated without breaching privacy issues. In operating outside privacy laws, behavioral data represents a new area of concern for scholars like Zuboff (2019), who term the free collection of this data "surveillance capitalism." In essence, for Zuboff (2019, p. 8), surveillance capitalism is "a new economic order that claims human experience as free raw material for…hidden commercial practices of extraction, prediction, and … behavioral modification." It is important to understand that surveillance capitalism, as facilitated by Big Data, is not a centralized state-controlled program that observes individual subjects but rather a market place, wherein behavioral data and "prediction products that anticipate what you will do now, soon, and later" are exchanged (Zuboff 2019, p. 14). Here, Zuboff's ethical issue with surveillance capitalism is *not* primarily related to privacy (as "individual users' meanings are of no interest to Google or other firms in this chain"), but rather with the "behavior modifications" these commodities promise (Zuboff 2015, p. 85). Surveillance in its market-form represents a profitable business model, wherein data and its modeling and algorithms are a sellable, revenue generating commodity traded on the promise of generating greater insights about the choices, behaviors, and activities of individuals in the future.

In the 2000s, surveillance capitalism evolved in concert with the growth of a 'data industry' that sat atop, and in part funded, the development of a new network infrastructure. A

vast array of databases were equipped to quickly cross reference internet user data stored in multiple locations through technologies like 'embedded cookies' (West 2017). During this time, Alphabet/Google, facing pressure from investors, made the pivotal decision to recruit user search histories and "click through rates" in developing an online advertising platform. The original service, known as AdWords, was possibly the progenitor of surveillance capitalism. The service was, in essence, an affiliation technology that tracked and recorded internet traffic through ads (Zuboff 2019). The business side of Google has since proliferated into a broad range of surveillance activities posing as free applications: Google maps track physical movements; Google Assistant and Google Home record personal and family interests and activities; and Google's AI messenger, Allo, tracks social communications. Together these platforms have allowed Google to amass an immense database of behavioral data, and in many cases this has also included personal data (Gaudin 2016; "Google tracks every move in location, maps even if you opt out," 2018; Hackett 2016). Google has also developed advanced "predictive analytics" or "artificial intelligence" that, together with user data, are sold as a complete advertising platform to those business customers wanting better access to, and influence over, their consumers. It is the costs associated with these new behavioral data practices, which we define as 'surveillance risk,' and it is the extraction, analysis and exchange of data within the broader system of surveillance capitalism that are of concern in this paper.

Prior work discussing the various ethical issues with surveillance has largely viewed it as part of a state apparatus, and focus on whether the use of big data sets by governments and agencies can be justified in terms of either national security (vis-à-vis terrorism) or public safety (vis-à-vis crime) (Etzioni 2018). Indeed, some have welcomed state surveillance, forwarding what is known as the "nothing to hide" argument, which posits that surveillance poses no threat to law-abiding individuals as government surveillance expressly targets unlawful activity (see Solove 2011). However, the evidence suggests that in order for surveillance activities to be effective they need to make similarly penetrating observations into the everyday lives of citizens, associated with or even just adjacent to suspects (Solove 2007, 2011). No doubt these peripheral forms of surveillance by the state are increasingly possible within the world of Big Data, as it stands replete with the technological capacity to search and aggregate information about individuals within it. The growth of a marketized form of surveillance no doubt intensifies these concerns as many governments and their agencies also participate in the exchange of behavioral data (Carr 2015; Lyon 2014; Zuboff 2019).

Other research has explored the role of surveillance as a hidden and possibly crucial aspect of the proper functioning of e-commerce platforms. For instance, Whelan (2019) argues that platforms like Uber work, not because of their ability to foster trust between strangers as Etzioni (2018) claims, but because of an underlying knowledge that their actions are observed (Rosenblatt 2017). Moreover, while these "strangers" are anonymous to each other on digital platforms, Whelan (2019, p. 16) argues that they are anything but anonymous to the platform itself, which requires private information such as "driver license information, bank account details, home address" as part of the enrolment process. In the same issue, Martin (2019) points out that choices around surveillance and privacy are entirely made by the website, platform, or 'market maker.' In this regard, the dramatic information asymmetries in big data represent a major shift in the power relationship between users and the platform, diminishing the ability of individual participants to hold internet companies accountable for mistreatment and other negative behavior (Rosenblat and Stark 2015).

The next section will further explore the underlying differences between privacy and surveillance risks and present data protection as a multi-dimension space of potential risk trade-offs.

A Data Protection Trade-Off: Privacy Risk Versus Surveillance Risks

In practice, concerns about both privacy and surveillance have been associated with the extraction, aggregation, and the trading of data on secondary markets that have rapidly expanded in scale and value over the last 5 to 10 years. While both concerns relate to possible harms that may be inflicted on individuals as a result of new forms of visibility, privacy and surveillance risks should not be conflated. There is little doubt that they are co-existing risks, but each produce different challenges to data regulators and society more broadly (Haggerty and Ericson 2000).

At its core, privacy deals with the rights of individuals to have control over the visibility of personal information such as our date of birth or our tax file number, but it also includes rights to our physical bodies, our emotions, personal relationships, our political views, and our choices. In conceptualizing privacy risks in this way, there has been a tendency to focus on the depth of penetration into aspects of personhood, rather than a consideration of these risks at scale. Privacy risks are thought to increase with the extraction of more detailed and more personal data because the depth of these data can be used to intimately target individuals as *subjects*. For the most part, privacy concerns are conceptualized at the scale of the individual and include a consideration of the implications of privacy breaches on each of us as individual members of society.

In contrast, the conceptualization of surveillance risk extends our understanding of these risks to include the broader control and governance implications of gathering and trading in this personal data. The scale, efficiency, and systematic nature of the acquisition of data, along with the inequities that exist between the organizations and institutions gathering the data and the data subjects, differentiates surveillance risk from privacy risk. Specifically, surveillance risks emerge because algorithms and analytics can be mobilized by large companies and governments to place individuals and groups in time and space in order to both understand past events and predict future behavior. Without doubt, there are a range of scenarios that include varying levels of both privacy and surveillance risks, but while data surveillance might benefit from access to personal information, it *does not need* this information to be effective. With access to anonymised data sets, it is possible to *infer* information about individuals and communities without the need for personal data or to make explicit links to individuals (Panackal and Pillai 2015; Soria-Comas and Domingo-Ferrer 2016; Xu et al. 2016).

As the accumulation of a wide variety of data has grown, new regulatory efforts have emerged to address both surveillance and privacy risks. However, these emergent forms of regulation have introduced new trade-offs between these risks, the effects of which we are yet to fully understand. In what remains of this paper, we interrogate these trade-offs, and we make the argument that the data protection options available under the EU's GDPR have not provided uniform improvements. The EU's effort to address privacy risk appears to have created space for new forms of surveillance. In the section that follows, we will unpack this in more detail, providing an analysis of the GDPR that highlights some of the new data-related challenges that are emerging in concert with the law.

The General Data Protection Regulation

Background

In contrast to the US, privacy regulators in Europe have been more successful in instituting restrictions on data collection and analysis (Safari 2016). Its latest law, the General Data Protection Regulation (EU) 2016/679 (GDPR) came into effect in 2018, superseding the 1995 Data Protection Directive as the overarching dictate on all European data. The GDPR aims to give citizens and residents of the European Union (EU) and the European Economic Area (EEA) control over their personal data, and to simplify the regulatory environment for international business by fully harmonizing the national data laws of its member states. The GDPR aims at creating a consistent, comprehensive approach to both privacy violations and mass surveillance for EU citizens.

While privacy or "Privacy by Design" (Art. 25) is the key concern for the GDPR, it is not limited to being simply a 'privacy act' per se. Instead, its objectives are inclusive and relate to the "fundamental rights and freedoms of natural persons" (Art. 1.1), surrounding the "processing" and "free movement of personal data" (Art. 1.1). A number of guidance and opinion documents were written by the Article 29 Working Party (the party responsible for drafting the law) relating directly to issues of surveillance with its aim to protect EU citizens from an array of Big Data-related methods and techniques levied both from within and by international companies and government agencies (WP 55, 215, 221, 228, 237[2]). No doubt, these working papers illustrate the overarching goal of the GDPR working party to restrict the "availability of large datasets and sophisticated analytics tools used to examine these datasets" by larger corporations who have a "dominant position" in Big Data market (WP 203, p. 46). In particular, a motivating factor in creating the GDPR was to address the rising power of these dominant Big Data players and address the "economic imbalance between [these companies] on one hand and consumers on the other" (WP 203, 2013, p. 46).

The GDPR has received praise as "a game changer" (Goodman 2018) and criticism as both a "property regime" (Victor 2013) and a stifling force for innovation and technological development (Zarsky 2016) due to its the extensive requirements in the processing of personal information by *data controllers* and *data processors*.[3] In contrast, and perhaps not surprisingly, the European Commission holds an optimistic view of the GDPR as "the gold standard" and "the most important change in data privacy regulation in 20 years" (eugdpr.org). The commission points to the enhanced rights, new consent practices, and heightened recourses and fines as key drivers contributing to greater organizational responsibility surrounding Big Data practices.[4] In comparison to the prior EU Data Protection Directive, the GDPR represents a significant intensification of the legal environment in that it is binding for all member states, and includes the rights to erasure and access; requirements for affirmative consent; and the ability to levy heavy fines. These fines can be in the order of €20 m or 4% of global revenues, whichever is higher, for especially severe violations (Art. 83.5). For less severe violations, Article 83.4 of

[2] We refer to these documents as "WP," see General Data Protection Regulation References for the full citations.

[3] *Data controllers* are those individuals or organizations responsible for personal data, and *data processors* are persons or organizations who use personal data as instructed by data controllers.

[4] https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf.

the GDPR sets forth fines of up to 10 million euros, or, up to 2% of its entire global turnover of the preceding fiscal year, whichever is higher. The first major filing in relation to the GDPR occurred in January 2019, wherein ten complaints have been filed by a European non-profit organization with the Austrian Data Protection Authority. The set of complaints represents an action of *€18.8 billion*, against eight online streaming companies including Apple and Amazon (CNBC 2019).

The Legal Compliance Options

We will first explore the main principles of the GDPR in cases where persons or data subjects are identifiable under the law. We will briefly explore how the law may offer an improvement to both privacy and surveillance risks. We then discuss the derogations available under the law in cases where de-identification techniques are used—this is our main concern. In short, many of the principles of the GDPR are not required when data controllers use data modifying 'safeguards.' Two main types of de-identification techniques are recognized under the law (anonymization and pseudonymization), both of which aim at either modifying or erasing personal data in order to protect the privacy of data subjects (see Hintze 2017; Hintze and El Emam 2018; Voigt and Von dem Bussche 2017). Our discussion section will then explore how these legal derogations and the use of de-identification techniques enhance surveillance risk for individuals.

Identified Data Sets

The GDPR assumes that, by default, data controllers wish to collect and process personal or identified data sets. In these instances, the GDPR requires that data controllers follow four principles that work together to form a data protection 'net' to safeguard their data subjects. The first two of these principles relate to *data collection*: data minimization and special categories. And the second two relate to *data processing*: purpose specification and automated decisions.

Data Collection In order to enact the principle of *data minimization,* the GDPR requires that the collection of data sets is "limited to what is necessary in relation to the purposes for which [it is] processed" (Art. 5.1.C). Indeed, the principle of data minimization is intuitive in that limiting the collection of the data set to "what is necessary" means that data controllers have limited options to undermine the data protection rights of their data subjects. Data are deemed to be personal if it relates to, or accurately describes, some aspect of a living, identifiable individual (Art. 4.1). While the collection of all personal data needs to be minimized under the GDPR "some forms of data are treated differently from others" (Zarsky 2016, p. 1012), here the EU data protection policy has a "layered regime," which recognizes that some forms of personal data are more sensitive than the other. Thus while data fields like *name, location, ID numbers, IP address*, and *economic information* are considered normal 'personal data' and should no doubt be limited under the data minimization principle, more sensitive data, known as *special categories,* are addressed by a separate article (Art. 9). Here, the GDPR "prohibits" the processing of what it deems to be sensitive data which "reveals racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life" except for in a narrow range of highly controlled circumstances that involve explicit consent (Art. 9.2a), reasons of substantial public interest or public safety (Art. 9.2i), or for special medical treatment (Art. 9.2 h).

Data Processing Data processing is dealt with in both the *purpose specification* and *automated decisions* principles. The former aims to provide subjects with more control over the movement of their personal information (Zarsky 2013). According to GDPR Article 5.1b purpose specification means that personal data can only be collected for a "specific, explicit and legitimate" purpose and not further "processed" (a term broadly defined) in ways that are not "compatible" with the original purpose. Another aim of the purpose specification principle is to ensure a temporal limitation to the use of data, so that "the period for which the personal data is stored, is limited to a strict minimum" and that, "time limits should be established by the controller for erasure or for a periodic review" (Rec. 39). As Zarsky (2016, p. 1006) argues, complying with the special purpose principle requires firms to "inform their data subjects of the future forms of processing they will engage in (which must still be legitimate by nature) and closely monitor their practices to assure they do not exceed the permitted realm of analyses." In principle, the GDPR also cautions against any unacknowledged use of *automated decisions* in the processing of data. The principle does not prohibit automation outright, but in an effort to preserve the rights of individuals subjected to these processes, it requires companies obtain explicit consent from individuals if they wish to use these techniques (Art. 22.2c). Moreover, in cases of automated decision making, organizations must provide "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject" if requested (Art. 13.2f).

The Privacy and Surveillance Risk of Identified Data Sets The above principles restrict activities that relate to both the privacy and surveillance risks, in that they impose significant restrictions and responsibilities onto data pro-

cessors, which stymie the flexibility of analysis, continued use, and distribution of personal data. In terms of privacy risk, the provisions set out within the GDPR represent a significant improvement to the EU's previous Data Protection Direction wherein many Big Data firms like Facebook were criticized for using disclosures as their main and sole form of responsibility regarding privacy (Martin 2016b, c). These practices came under fire in a number of European court cases (e.g., Shrems vs. Irish Data Protection Commissioner) for heaping privacy management upon the individual and "reinforcing the idea that we are to treat ourselves as if we were an enterprise that should undertake risk management" (Hull 2015; Richardson 2016, p. 69). No doubt, the principles of data minimization and special categories enforce a new standard in data collection, which alleviates the data subject of the individual burden of having to understand, manage, and monitor his/her own data. These principles cannot be circumvented through crafty, self-serving consent practices, disclosures, or contractual arrangements that limit a data controller's responsibility. In addition, the GDPR further prevents the short-term trading of personal data for benefits and access to services through new data rights, such as the right to be forgotten (Art. 17.2) and the right to be informed (Art. 13).

The adherence to strict parameters and practices of a predefined and limited scope (purpose specification) prevents data controllers from expanding their operations and selling the information for any incompatible use. In introducing the principle of purpose specification, the potential lifespan and proliferation of the data for subsequent surveillance by companies and government agencies are somewhat limited. Any variation to the stated purpose, which may include the intention to sell the data, requires additional consent processes that must include an opportunity to opt-out (Art. 6). Here the purpose specification aims to curtail scope drift or "inhibit 'mission creep', which could otherwise give rise to the usage of the available personal data beyond the purposes for which they were initially collected" (WP 203, p. 4). Moreover, the minimization principle also prevents data controllers from stockpiling additional "just-in-case" information, extraneous to a specific purpose, in the anticipation that this information may be of value in the future. This significantly reduces the value of personal (identifiable) data sets for data brokers as they are less transferable to businesses that have other data processing purposes.

Despite the above principles and new restrictions present when using identifiable data sets, a level of privacy risk is still present for individuals because the law allows companies to store personal data with a significant level of detail as long as the purpose is specified and the individual is informed. This is logical given the many situations in which our personal identity is relevant, such as a record with our health provider or mortgage lender. The GDPR principles have also introduced some additional protections against the potential use of these data for surveillance purposes. Specifically, the principles of *purpose specification* and *data minimization* limit the scope of use and free exchange of all data collected by controllers.

De-identified Data Sets

We will now explore the treatment of de-identified data sets under the GDPR. Here the law allows data controllers to choose between two different approaches to manage this data: a*nonymization* or *pseudonymization*. In recognizing the privacy-enhancing effect of these additional safeguards, the GDPR provides exceptions to many of the most burdensome provisions of the regulation, particularly in terms of *purpose specification*. We will explore how, rather than being an outright improvement to data protection, anonymization and pseudonymization techniques create different trade-offs in relation to the data protection they offer.

Anonymized Data Recital 26 of the GDPR defines anonymized data as a viable safeguard wherein "data [is] rendered anonymous in such a way that the data subject is not or no longer identifiable" by "all the means reasonable" and that this modification must be irreversible (Directive 95/46/EC). As a consequence, data that are anonymized within the GDPR have all personally identifiable material deleted from the data set. Under the law, the defining aspect of this 'irreversibility' is whether a controller or third party can show they have used "all the means likely" and "reasonably" to prevent de-anonymization (Rec. 26). While the EU does not provide a standard of successful anonymization per se, it advises using a combination of randomization and generalization techniques in order to create stronger privacy guarantees. For companies facing the new restrictions of the GDPR, anonymization is an appealing option, not only in its protection of privacy, but because, in anonymizing, all the other elements of the law become redundant. Recital 26 GDPR clearly states that the "regulation does not … concern the processing of … anonymous information," thus, data controllers and processors do not have to abide by the above four principles of data protection. However, in order to comply, anonymization requires "careful engineering" and "constant monitoring" in order to ensure the data remain de-identified (Voigt and Von dem Bussche 2017, p. 15).

Pseudonymized Data While anonymization is an option within the GDPR, the regulation suggests a clear preference for de-identification through 'pseudonymization.' Article 4.5 of the GDPR defines pseudonymization as "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information." The GDPR believes pseudonymi-

zation meets "the principles of data protection by design and … default" (Rec. 78). Under the law, the difference between *pseudonymization* and *anonymization* rests on whether the data can be re-identified. Unlike anonymization, if data are pseudonymized, the additional personal information is not destroyed, but is held separately and securely, through the use of "technical and organizational measures" to "ensure that the personal data is not attributed to an identified or identifiable natural person" (Art. 4.5). No doubt, there is a risk that the data can become identifiable if both the "personal" and "additional" information elements are brought together. In response to these concerns, and unlike anonymization, the GDPR states that both the personal information and de-identified information fall within the law; however, the use of pseudonymization allows for a relaxation of the purpose specification principle. For instance, Article 6(4)(e) permits the processing of pseudonymized data for uses *beyond the purpose for which the data were originally collected*. This "general analysis" also includes the freedom to sell de-identified data for similarly general (i.e., unrestricted) use (Rec. 29) in order to "create incentives" for controllers and processors "to apply pseudonymization." However, data controllers are still required to fully comply with other aspects of the GDPR, such as requirements to gather permissions and consent from data subjects, as well as the need to provide guarantees and disclosures of data policies. The relaxation of the purpose specification principle means that pseudonymization is an attractive option for controllers, allowing for regulatory compliance with the GDPR, while permitting an expansion of data collection and the subsequent uses of this data. Pseudonymization might prove to be the best option for those data controllers who, after assessing both their own interests and those of their data subjects, need greater regulatory flexibility while still controlling and storing personal information.

The privacy and surveillance risk of de-identified data sets: Anonymization and pseudonymization techniques represent either a deletion or hiding of personal data and thus significantly reduce privacy risk. As a result, these techniques may seem, at least on the surface, to be an overall improvement in data protection. However, whether the general data protection regulation and its stance on anonymization and pseudonymization have come about through its emphasis on individual privacy, an overzealous trust in technological solutions to data risk, it is clear that the law, rather than generate and improve the state of data protection outright has created different trade-offs related to risk. In the following discussion, we will argue that the law lacks a vital understanding of contemporary big data practices and that improvements in terms of privacy also bring underlying and systematic expansion of surveillance risk. While pseudonymized data provide certain additional protections for individual privacy, it introduces higher surveillance risk than situations in which no identification techniques are used and where data controllers are required to observe the principles of purpose specification and data minimization. Perhaps more crucially, in practice, the line is blurred between pseudonymization and anonymization; here the former can become the latter through the deletion of the associated identification data (Mourby et al. 2018). In the following section, we will then discuss some of the wider consequences of the GDPR's de-identification techniques, paying particular attention to the growing market for de-identified data and its broader implications.

De-identification Techniques in the Age of Surveillance Capitalism

While the use of the above data modification techniques is currently *legal* the remainder of this paper will now consider the ethical implications of their use, and the possible underlying effects on surveillance risk. We will draw on the insights from Zuboff's (2019) book titled *The Age of Surveillance Capitalism* and introduce the idea of de-identified data as *res nullius* in order to explain our concerns. This discussion will explore the possible ways the GDPR may be further supporting a new type of 'market-based' surveillance, which emerged in the last 15 years and which represents new harms to individuals that are only beginning to be understood.

Zuboff's (2019) *The Age of Surveillance Capitalism* details and traces the activities of a new breed of data capitalist; one driven to acquire ever more predictive sources of behavioral data on our choices, activities, behaviors, personalities, and emotions. While, originally pioneered by Google, the surveillance business model and the collection of customer data are now de rigueur for Chief Information Officers, not wanting to miss out on the "Big Data Gold Rush" (Peters 2012). While the extraction of valuable resources by corporate actors is by no means new, what is novel under the current epoch of surveillance capitalism is the unilateral "claims [made of] human experience as free raw materials for translation into behavioral data" (Zuboff 2019, p. 8). The economic value of this "behavioral surplus"[5] is then further enhanced when combined with "machine intelligence," such as algorithms, which generate predictive modeling of human activities (Zuboff 2019, p. 74). Behavioral data is now collected by a growing number of devices and applications and, while the general public may be aware that Alexa, Google Home, and Google Maps collect their personal data, there is less consciousness of the ongoing extraction of deep and detailed behavioral patterns.

[5] The collection and trade of behavioral data, above the company's own internal requirements for product and service improvements.

Moreover, there is a proliferation of new platforms, such as games (like Pokemon Go), wearable devices (such as Fitbits), and even children's toys that have the ability to 'spy,' that is, capture and transmit other forms of data, such as our children's first words (Federal Trade Commission 2016).

For Zuboff (2019), the system of 'surveillance capitalism' does not operate as a centralized panopticon that identifies and targets individuals, but as a de-centralized marketplace that exchanges both behavioral data (as a raw commodity), and its associated predictive models to analyze this data. The economic values of these goods and services are determined by *what* they can do, not *who* they are about. In fact, "individual users' meanings are of no interests to Google" or other big data companies, but rather the "behavioral implications" they promise (Zuboff 2015, p. 79). In this revision, rather than being strictly a function of the state, surveillance is an evolved capacity of capitalism itself as its key players re-shape various technologies for their own profitable ends (Zuboff 2019).[6] Within this new surveillance-based business model, behavioral data is treated as a free commodity—a *res nullius*[7]—largely ungoverned by legal frameworks. In short, behavioral data is an object with no owner. Here data subjects, anonymized or otherwise, are afforded no ownership rights of their own behavioral data. Zuboff (2019, p. 377) draws the analogy of the role of elephants within the Ivory Trade. Here, we, like these "Majestic mammals," are not the product but "the abandoned carcass"; the product on the other hand has been "ripped from [our] lives." This decoupling of behavioral data from its subjects is a new technological form of alienation.

An ethical analysis of the GDPR and, in particular, the derogations it affords to the use of de-identifying techniques, must be made in relation to these marketized forms of surveillance. While the law was designed to protect individuals from surveillance risks (WP 55, 215, 221, 228, 237) it may be, in actuality, intensifying these issues. Here, the treatment of anonymized data as outside the scope of the law and, to a lesser extent, the relaxation of the *purpose principle* in the case of pseudonymized data are indicative of what the law sees as the main target of big data practices—personal information. At its core, the aim of the law is to reduce harmful forms of identification (preamble, paragraph 26) and thus de-identification techniques respond to this aim through preventing the "singling out" of "natural persons." In fact, the robustness of anonymization and pseudonymization is measured by how well subjects are separated and guarded from the data they produce (Rec. 29; WP 216, p. 3). Here, the GDPR's focus on individual privacy may be unintentionally providing data controllers with a legal mandate to pursue market surveillance practices as they attempt to avoid the costly constraints and procedures required by the regulation. Thus, paradoxically the same de-identification 'safeguards,' designed to shield individuals from the extraction of their private data, may be facilitating, and even institutionalizing, the exchange of their behavioral data. At present, the GDPR does not recognize the potential harms arising from the behavioral data. Rather its concern for "monitoring the behavior of … data subjects" again relates to whether the data can be clearly tied to a "natural person" and "personal data" (Rec. 24). Moreover, the law, in approving the use of complex modification techniques like anonymization and pseudonymization may be inadvertently crystallizing the power of tech elites, like Google and Facebook, which have already established vast economies of scale in the collection and analysis of behavioral data. This is the very opposite of the stated objectives of the law, which are to create a more competitive field and to reduce the power of those companies with a "dominant position" (WP 203 2013, p. 46). Rather, the legal treatment of de-identification may directly increase the *scale of visibility,* with ever increasing numbers of individuals and spaces being subject to the gaze of a few powerful companies.

Perhaps most importantly, Zuboff's (2019) key concern with surveillance capitalism is not processes that seek "not only to know [us]" but rather those that seek to "shape our behavior at scale," that is to enhance social control. Here the competitive pressures to maximize profits have led to a shift in the way that behavioral data is used; "the goal now is to automate us" through "complex and comprehensive means of behavior modification" (Zuboff 2019, p. 15). Through suturing predictive algorithms to 'tuning,' 'herding,' and 'conditioning techniques,' web platforms have managed to change both online and offline behavior in ways that are beneficial to their own goals (Lyon 2014; Zuboff 2019). For instance, in a 2014 experiment, Facebook was not only able to directly influence the views of its users through changing its newsfeed, but found that these 'manipulations' also affected the views of friends and family connected within the same local network (Kramer et al. 2014). Thus, the value of the advertising services of tech platforms like Facebook is the access to data, predictive algorithms, and behavioral modification techniques, the valuable combination of which has been described as an emerging "behavioral futures market" by Zuboff (2019, p. 15). The findings of Kramer et al. (2014) no doubt reiterate concerns surrounding this powerful combination of technologies to influence the preferences and emotions of consumers and voters (Lyon 2003, p. 14, 2014). The potential for reconstructing the relationship between individuals, companies, and states at a fundamental level is massive here, and is what Lyon (2014, 2015) refers

[6] While the state is still a major facilitator and participant in this surveillance model (West 2017) it is no longer the central player.

[7] A Latin term that means an object or property outside the legal rights framework.

to as "anticipatory governance," where Big Data is used to both measure and influence popular thinking around matters of public concern. The sanctioned use of behavioral data via the GDPR thus has the potential to play a new central role in the management of the domestic political economy (Kitchin 2014, p. 165) and lead to what Zuboff (2015) refers to as a loss of "reciprocity" between the ruling elites and the population, increasingly usurping the ability of individuals to make decisions. If we accept that liberal democratic societies are subject to a valuable measure of "uncertainty" that has forced governments and companies to listen and adapt to the needs of its citizens and customers (Zuboff 2015), then Big Data may represent a break with this past. Here, the side effect of the development of the behavioral data ecosystem is an increasing ability to mitigate 'the unknown' that both challenge power relations and form the basis of a vibrant democracy. This is increasingly being replaced by 'the known' and, under the GDPR, will be progressively more tradeable in exclusive behavioral futures markets (Zuboff 2019).

In the following section, we propose some recommendations to both acknowledge and balance some of the above issues with the GDPR.

Recommendations

There are two clear recommendations that emerge from our analysis. The first relates to the need for a proper legal consideration of the property rights surrounding the ownership of behavioral data; what we have referred to as an object with no owner (*res nullius)*. Such a consideration may challenge the assumption of the GDPR that techniques such as anonymization and pseudonymization separate data from individuals. In order to future proof citizens from the wider implications of data collection and trade, we recommend that regulators widen their net. While the secure and transparent management of personal identifiers within data is critical, the growing market for pseudonymized behavioral data remains outside the purview of the current law. Moreover, new laws surrounding the ownership status of behavioral data need to consider the broader context in which behavioral data operates, complete with its capabilities to make accurate predictions and inferences about people. No doubt this will require a greater philosophical engagement with not only what behavioral data *is*, but also the deeper question as to whether an individual's behaviors, actions, and emotions can be considered to be their own property, and if so, whether the recording, measuring, or trading of such phenomena is an infringement of individual rights. These questions remain important, even where the owner of said property is anonymized. No doubt, broader debates about the ownership of behavioral data have already started to take place, and while outside the scope of this paper, they represent important areas of future research (Miller 2014, p. 135; Xiaying 2019). Our recommendation is not necessarily wedded to the establishment of new property rights over *res nullius* data but may encompass alternative approaches, such as a stakeholder perspective that would ask the question: 'if Big Data can predict and modify the activities of people with great accuracy then should it not stand to reason these same individuals have a stake in its function?'

Our second recommendation is more specific. For the GDPR to be robust, data that are collected with the intention to be anonymized should still be subject to the consent and agreement of the individual. While pseudonymization requires consent processes, under the current law, anonymized data do not. If we view consent statements as solely legal contracts between parties, then the EU Commission's decision to exempt anonymized data from this process makes sense. Indeed, there are obvious problems with trying to establish an agreement between a data controller and an anonymous party. However, consent processes have a number of other important functions, such as disclosing data protection policies and rights, as well as obtaining an acknowledgement of these rights from data subjects (Hintze 2016). From this perspective, it is important to ensure all individuals know if their data can be traded, even if the data are completely anonymized, and to grant them the right not to make their anonymous data available for this purpose.

Conclusion

The era of Big Data has only just begun, however, understanding its risk of harm to individuals is still of utmost importance. This is particularly the case as governments continue to shape legal, regulatory, and conceptual frameworks aimed at stemming these risks, while also allowing for a safe use of Big Data. Research continues to explore whether and how the GDPR strikes a balance between the free use of data and the protection of citizens (Zarsky 2016). In order to contribute to this discussion, our paper has explored the tensions arising within the GDPR, as it seeks to satisfy both data-related privacy and surveillance concerns. We argue that the current framing of the law has made significant inroads in terms of protecting individual privacy, whether identified or de-identified. However, this emphasis placed on individual privacy does not provide sufficient protections to citizens against the surveillance risks associated with the collection, analysis, and trade of their behavioral data. Indeed, the way the law has been constructed has created space for a behavioral futures market to emerge with very few protections for individuals. Moreover, and in concert with Zuboff (2019), we argue that the rise of surveillance practices is changing the very fabric of society.

So, in protecting the identity of the individual, our paper argues that the law must also protect the increasingly visible behavioral aspects of people as their political, social, and economic choices constitute the future.

There is little doubt that the GDPR represents a bold attempt by the EU to introduce new data protection regulations, and it will go some way towards addressing emergent ethical concerns about data collection and processing. However, its reach will be limited if law makers fail to understand the broader behavioral data ecosystems they seek to regulate. Given this, new research that attempts to theorize and empirically ground our understanding of 'surveillance capitalism' will be crucial (see Lyon 2003, 2014; Lyon et al. 2012; West 2017; Zuboff 2015, 2019). Based on our reading of the law, at present the GDPR's effort to codify data ethics, particularly in relation to behavioral data, is limited because its derogations create a passage through which companies are able to escape the law's restrictions. Indeed, the law creates the space for a behavioral data market in which commercial self-interest is likely to flourish. In this way, whether deliberately or otherwise, the GDPR will make it possible for the behavioral data market to continue to function unencumbered. It is here that we have significant concerns because both the nature of the data market and the trading of this kind of data are open to new strategies that are being deployed to modify behavior (Zuboff 2019). As it currently stands, the GDPR is likely to make data controllers more aware of their responsibilities to data subjects, but it provides little protections against the misuse of data within these markets. The management and oversight of Big Data will continue to produce significant ethical dilemmas as regulators try to keep pace with the needs of the community and the commercial imperatives of an increasingly lucrative market. If regulators are to do this well, the surveillance implications of big data will need to be taken just as seriously as our well-founded privacy concerns.

Compliance with Ethical Standards

Research Involving Human and Animal Participants This article does not contain any studies with human participants or animals performed by any of the authors.

General Data Protection Regulation References

Art. 29 WP 55 (2002). *Working document on the surveillance of electronic communications in the workplace* (No. 5401/01/EN/Final). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 203 (2013). *Opinion 03/2013 on purpose limitation* (No. 00569/13/EN). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 215 (2014). *Opinion 04/2014 on surveillance of electronic communications for intelligence and national security purposes* (No. 819/14/EN WP 215). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 216 (2014). *Opinion 05/2014 on anonymization techniques* (No. 0829/14/EN). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 221 (2014). *Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU* (No. 14/EN WP 221). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 228 (2014). *Working document on surveillance of electronic communications for intelligence and national security purposes* (No. 14/EN WP 228). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 237 (2016). *Working document 01/2016 on the justification of interferences with the fundamental rights to privacy and data protection through surveillance measures when transferring personal data (European Essential Guarantees)* (No. 16/EN WP 237). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

Art. 29 WP 251 (2017). *Guidelines on automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (No. 17/EN WP 251). Brussels, Belgium: Article 29 Data Protection Working Party, European Commission, Directorate General Justice.

References

Andrejevic, M. (2014). Big data, big questions: The big data divide. *International Journal of Communication, 8,* 17.

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society, 12*(2), 185.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. *Privacy, Big Data, and the Public Good: Frameworks for Engagement, 1,* 44–75.

Baruh, L., & Popescu, M. (2017). Big data analytics and the limits of privacy self-management. *New Media & Society, 19*(4), 579–596.

Berk, R. (2012). *Criminal justice forecasts of risk: A machine learning approach*. New York: Springer.

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science, 350*(6264), 1073–1076.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679.

Calmon, F. P., Makhdoumi, A., & Médard, M. (2015). Fundamental limits of perfect privacy. In *June 2015 IEEE international symposium on information theory (ISIT)* (pp. 1796-1800). IEEE.

Carr, M. (2015). Power plays in global internet governance. *Millennium, 43*(2), 640–659.

Chaudhuri, A. (2016). Internet of things data protection and privacy in the era of the General Data Protection Regulation. *Journal of Data Protection & Privacy, 1*(1), 64–75.

Cooley, T. M. (Ed.). (1888). *A treatise on the law of torts, or the wrongs which arise independent of contract*. 2nd edn. Chicago: Callaghan & Co.

CNBC. (2019). Austrian data privacy activist files complaint against Apple, Amazon, others. Retrieved January 18, 2019, from https://www.cnbc.com/2019/01/18/austrian-data-privacy-activist-files-complaint-against-apple-amazon-others.html.

Englehardt, S., & Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1388–1401). ACM.

Etzioni, A. (2018). Apple: Good business, poor citizen? *Journal of Business Ethics, 151*(1), 1–11.

Federal Trade Commission. (2016). In the matter of genesis toys and nuance communications—Complaint and request for investigation. Submitted by The Electronic Privacy Information Center, The Campaign for a Commercial Free Childhood and The Center for Digital Democracy Consumers Union, 6 December, 2016.

Floridi, L. (2014). Open data, data protection, and group privacy. *Philosophy & Technology, 27*(1), 1–3.

Gandy, O. H., Jr. (2012). Statistical surveillance: Remote sensing in the digital age. In K. Ball, K. Haggerty, & D. Lyon (Eds.), *Routledge Handbook of surveillance studies* (pp. 125–132). London: Routledge.

Gaudin, S. (2016). How Google Home's "always on" will affect privacy. Computerworld. Retrieved October 6, 2016, from https://www.computerworld.com/article/3128791/how-google-homes-always-on-will-affect-privacy.html.

Geng, H. (2017). *Internet of things and data analytics handbook*. Chichester: Wiley.

Glaeser, E. L., Hillis, A., Kominers, S. D., & Luca, M. (2016). Crowdsourcing city government: Using tournaments to improve inspection accuracy. *American Economic Review, 106*(5), 114–118.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry, 56*(1), 114–137.

Goodman, S. (2018). A game changer in the personal data protection in the EU. *MSU Internatonal Law Review*. Retrieved from https://www.msuilr.org/msuilr-legalforum-blogs/2018/2/19/a-game-changer-in-the-personal-data-protection-in-the-eu. Retrieved 23 Aug 2018.

Google tracks every move in location, maps even if you opt out. (2018). News Corp. Retrieved August 14, 2018, from https://www.news.com.au/technology/gadgets/mobile-phones/google-has-been-tracking-your-movements-even-if-you-told-it-not-to/news-story/bb9eb906387ffd2295e8b17b24b7d883.

Gritzalis, S. (2004). Enhancing web privacy and anonymity in the digital era. *Information Management & Computer Security, 12*(3), 255–287.

Hackett, R. (2016). Everything you need to know about Google Allo's Privacy Backlash. Fortune. Retrieved September 22, 2016, from http://fortune.com/2016/09/22/google-allo-nope/.

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology, 51*(4), 605–622.

Hintze, M. (2016). In defense of the long privacy statement. *Md. L. Rev., 76,* 1044.

Hintze, M. (2017). Viewing the GDPR through a de-identification lens: A tool for compliance, clarification, and consistency. *International Data Privacy Law, 8*(1), 86–101.

Hintze, M., & El Emam, K. (2018). Comparing the benefits of pseudonymisation and anonymisation under the GDPR. *Journal of Data Protection & Privacy, 2*(2), 145–158.

Hull, G. (2015). Successful failure: What Foucault can teach us about privacy self-management in a world of Facebook and big data. *Ethics and Information Technology, 17*(2), 89–101.

Kamara, I. (2017). Co-regulation in EU personal data protection: The case of technical standards and the privacy by design standardisation'mandate'. *European Journal of Law and Technology*, *8*(1). http://ejlt.org/article/view/545/723.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Thousand Oaks: Sage.

Kokolakis, S. (2017). Privacy attitudes and privacy behavior: A review of current research on the privacy paradox phenomenon. *Computers & Security, 64,* 122–134.

Koops, B.-J., & Leenes, R. (2014). Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design'provision in data-protection law. *International Review of Law, Computers & Technology, 28*(2), 159–171.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788–8790.

Lohr, S. (2012). The age of big data. *New York Times*. Retrieved February 11, 2012, from https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

Lyon, D. (2003). *Surveillance as social sorting: Privacy, risk, and digital discrimination*. London: Psychology Press.

Lyon, D. (2014). Surveillance, snowden, and big data: Capacities, consequences, critique. *Big Data & Society, 1*(2), 2053951714541861.

Lyon, D. (2015). *Surveillance after snowden*. New York: Wiley.

Lyon, D., Haggerty, K. D., & Ball, K. (Eds.). (2012). *Routledge handbook of surveillance studies*. New York: Routledge.

Madden, M. (2014). *Public perceptions of privacy and security in the post-Snowden era*. The Pew Research Centre. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/14/2014/11/PI_PublicPerceptionsofPrivacy_111214.pdf. Retrieved 20 Feb 2018.

Madden, M., Gilman, M., Levy, K., & Marwick, A. (2017). Privacy, poverty, and big data: A matrix of vulnerabilities for poor Americans. *Washington University Law Review, 95,* 53–125.

Martin, K. (2015a). Ethical issues in the big data industry. *MIS Quarterly Executive, 14*(2), 67–85.

Martin, K. (2015b). Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing, 34*(2), 210–227.

Martin, K. (2016a). Data aggregators, consumer data, and responsibility online: Who is tracking consumers online and should they stop? *The Information Society, 32*(1), 51–63.

Martin, K. (2016b). Do privacy notices matter? Comparing the impact of violating formal privacy notices and informal privacy norms on consumer trust online. *The Journal of Legal Studies, 45*(S2), S191–S215.

Martin, K. (2016c). Understanding privacy online: Development of a social contract approach to privacy. *Journal of Business Ethics, 137*(3), 551–569.

Martin, K. (2019). Trust and the online market maker: A comment on Etzioni's Cyber Trust. *Journal of Business Ethics, 156*(1), 21–24.

Martin, K., & Nissenbaum, H. (2016). Measuring privacy: An empirical test using context to expose confounding variables. *Columbia Science and Technology Law Review, 18,* 176–218.

Martin, K., & Shilton, K. (2016). Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society, 32*(3), 200–216.

Marwick, A. E. (2012). The public domain: Social surveillance in everyday life. *Surveillance & Society, 9*(4), 378–393.

Miller, K. (2014). Total surveillance, big data, and predictive crime technology: Privacy's perfect storm. *Journal of Technology Law & Policy, 19,* 105–146.

Miraz, M. H., Ali, M., Excell, P. S., & Picking, R. (2015). A review on Internet of Things (IoT), Internet of everything (IoE) and Internet of nano things (IoNT). In *Internet technologies and applications (ITA), 2015* (pp. 219–224). IEEE.

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., et al. (2018). Are 'pseudonymised'data always personal data?

Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review, 34*(2), 222–233.
Newkirk, D. (2018). "Apple: good business, poor citizen": A practitioner's response. *Journal of Business Ethics, 151*(1), 13–16.
Newman, N. (2013). The costs of lost privacy: Consumer harm and rising economic inequality in the age of Google. *William Mitchell Law Review, 40,* 849–889.
Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Standford, CA: Stanford University Press.
Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *The New England Journal of Medicine, 375*(13), 1216–1219.
Pagallo, U. (2017). The legal challenges of big data: Putting secondary rules first in the field of EU data protection. *European Data Protection Law Review, 3,* 36–46.
Panackal, J. J., & Pillai, A. S. (2015). Adaptive utility-based anonymization model: Performance evaluation on big data sets. *Procedia Computer Science, 50,* 347–352.
Peters, B. (2012). The big data gold rush. Forbes. Retrieved June 21, 2012, from https://www.forbes.com/sites/bradpeters/2012/06/21/the-big-data-gold-rush/.
Qualman, E. (2010). *Socialnomics: How social media transforms the way we live and do business*. New York: Wiley.
Richardson, J. (2016). *Law and the philosophy of privacy*. Oxford: Routledge.
Rosenblatt, A. (2017). How drivers shame Uber Lyft passengers. Medium, May 29, 2017.
Rosenblat, A., & Stark, L. (2015). *Uber's drivers: Information asymmetries and control in dynamic work*. Data & Society Research Institute, 17.
Rotenberg, M., & Jacobs, D. (2013). Updating the law of information privacy: The new Framework of the European union. *Harvard Journal of Law & Public Policy, 36,* 605–652.
Rouvroy, A. (2016). "Of Data and Men". Fundamental rights and freedoms in a world of big data. *Council of Europe, Directorate General of Human Rights and Rule of Law, T-PD-BUR (2015) 09REV,* Strasbourg.
Safari, B. A. (2016). Intangible privacy rights: How Europe's GDPR will set a new global standard for personal data protection. *Seton Hall Law Review, 47,* 809–848.
Scott, M. (2018). Cambridge Analytica helped 'cheat' Brexit vote and US election, claims whistleblower. Retrieved September 10, 2018, from https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/.
Solove, D. J. (2007). Nothing to hide. *San Diego Law Review, 44*(4), 745–772.
Solove, D. J. (2011). *Nothing to hide: The false tradeoff between privacy and security*. London: Yale University Press.
Soria-Comas, J., & Domingo-Ferrer, J. (2016). Big data privacy: Challenges to privacy principles and models. *Data Science and Engineering, 1*(1), 21–28.
Taddicken, M. (2014). The 'privacy paradox'in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication, 19*(2), 248–273.
Taylor, L., Floridi, L., & Van der Sloot, B. (2016). *Group privacy: New challenges of data technologies* (Vol. 126)., Philosophical Study Series New York: Springer.
Torra, V. (2017). *Data privacy: Foundations, new developments and the big data challenge*. Cham: Springer.
Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197–208.
Victor, J. (2013). The EU General Data Protection Regulation: Toward a property regime for protecting data privacy. *The Yale Law Journal, 123*(2), 513–528.
Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR)* (Vol. 18). Cham: Springer.
Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review, 34*(3), 436–449.
Warner, R., & Sloan, R. H. (2014). Self, privacy, and power: Is it all over. *Tulane Journal of Technology and Intellectual Property, 17,* 61–108.
Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review, 4*(5), 193–220. https://doi.org/10.2307/1321160.
West, S. M. (2017). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society, 58,* 1–22.
Whelan, G. (2019). Trust in surveillance: A reply to Etzioni. *Journal of Business Ethics, 156*(1), 15–19.
Xiaying, M. (2019). The legal attributes of electronic data and the positioning of data in civil law. *Social Sciences in China, 40*(1), 82–99.
Xu, L., Jiang, C., Chen, Y., Wang, J., & Ren, Y. (2016). A framework for categorizing and applying privacy-preservation techniques in big data mining. *Computer, 49*(2), 54–62.
Young, A. L., & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The Internet privacy paradox revisited. *Information, Communication & Society, 16*(4), 479–500.
Zarsky, T. Z. (2013). Transparent predictions. *University of Illinois Law Review, 2013,* 1503–1569.
Zarsky, T. Z. (2016). Incompatible: The GDPR in the age of big data. *Seton Hall Law Review, 47,* 995–1020.
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*(1), 75–89.
Zuboff, S. (2019). *The age of surveillance capitalism: the fight for the future at the new frontier of power*. New York: Profile Books.