Routledge
Taylor & Francis Group

The Information Society
An International Journal

ISSN: 0197-2243 (Print) 1087-6537 (Online) Journal homepage: http://www.tandfonline.com/loi/utis20

Big data privacy: The datafication of personal information

Jens-Erik Mai

To cite this article: Jens-Erik Mai (2016) Big data privacy: The datafication of personal information, The Information Society, 32:3, 192-199

To link to this article: http://dx.doi.org/10.1080/01972243.2016.1153010

THE INFORMATION SOCIETY
2016, VOL. 32, NO. 3, 192–199
http://dx.doi.org/10.1080/01972243.2016.1153010

Routledge
Taylor & Francis Group

Big data privacy: The datafication of personal information

Jens-Erik Mai

Royal School of Library and Information Science, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

In the age of big data we need to think differently about privacy. We need to shift our thinking from definitions of privacy (characteristics of privacy) to models of privacy (how privacy works). Moreover, in addition to the existing models of privacy—the surveillance model and capture model—we need to also consider a new model: the datafication model presented in this article, wherein new personal information is deduced by employing predictive analytics on already-gathered data. These three models of privacy supplement each other; they are not competing understandings of privacy. This broadened approach will take our thinking beyond current preoccupation with whether or not individuals' consent was secured for data collection to privacy issues arising from the development of new information on individuals' likely behavior through analysis of already collected data—this new information can violate privacy but does not call for consent.

ARTICLE HISTORY
Received 10 December 2014
Accepted 24 January 2016

KEYWORDS
Big data; datafication; personal information; privacy

Predictive analytics was not widely known until Duhigg (2012) in the winter of 2012 wrote a small piece in *The New York Times* in which he told the story of a father who went to a Target store complaining that his daughter had received coupons for maternity clothing and baby products. It turned out that Target knew before anyone else that the daughter was pregnant. Target had collected data about the daughter's purchase history for some 25 unique products, which, when analyzed together, produced a "pregnancy prediction" score. Big data and predictive analytics all of a sudden became very concrete for the public—and people came to realize that personal information is in fact a commodity that is sold and traded among information empires and data brokers.

Whether or not the daughter's privacy had been breached is a complicated question. It appears that the daughter had volunteered information about her purchases to Target; she had the ability to limit access to the information and she did control the flow of the information. However, she might not have realized that Target could pry into her personal life by datafying her personal information and creating new information about her that she had not volunteered to Target. While she did control the information she provided to Target, she had no control over the knowledge that Target produced based on the information she had provided to Target.

This article introduces a datafication model of informational privacy. The basic idea is that as big data becomes mainstream and businesses and state agencies apply predictive analysis to generate new information and knowledge about customers and citizens, a shift in focus from data collection to data processing is needed. In the age of big data we need be concerned not only about the collection of data but equally about the processing of data to generate new information and knowledge. The datafication model of privacy introduced in this article helps address this challenge.

This article is organized as follows: It first discusses the notion of privacy in the age of big data. The article then discusses five conceptual challenges associated with the notion of informational privacy: (i) the notion of information, (ii) the focus on individual pieces of information, (iii) the subjectivity of privacy, (iv) the distinction between private and public, and (v) the value of privacy. Thereafter, the article augments Agre's (1994) two models of privacy—sthe urveillance model and capture model—with a third model: a datafication model of informational privacy. Lastly, the article shows how the five challenges to privacy are being reconfigured in the age of big data privacy.

Big data privacy

People reveal personal information consciously or unconsciously, willingly or unwillingly, as they perform everyday activities: shopping for groceries, communicating with family members, paying taxes, reading the news,

CONTACT Jens-Erik Mai jemia@hum.ku.dk Royal School of Library and Information Science, University of Copenhagen, Birketinget 6, 2300 Copenhagen, Denmark.

listening to music, reading e-books, purchasing gasoline, exchanging e-mails, sharing photos, and so on. In addition, many people choose to reveal information about their private lives on social networking sites. While the latter may be called the "great privacy give-away" (Allen, 2013, 847) or "media exhibitionism" (Nissenbaum 2010, 106), this article does not make distinctions between reasons for revealing information but merely notes that it is almost impossible to perform most daily activities without revealing personal information and providing fodder for data brokers and big data organizations, whether they are private or public. Given that this personal information is often revealed on a voluntary basis, as people willingly provide personal information while interacting with digital entities—in some cases they might even explicitly have consented to the organizations collecting their personal information—it could reasonably be argued that people have relinquished their right to privacy with regard to that information. Assuming that this is right, that people have willingly provided their personal information in their daily activities, we should question whether they still have a moral right not to have that information applied in predictive analytics to gain insights into their personal lives and preferences. To address that question, we might have to reconceptualize the notion of privacy to fit the current big data paradigm.

Solove (2013) notes that "at the core of many privacy issues is the consent dilemma" (1903), and much privacy practice and theory relies on the notion that people have the right or opportunity to consent to provide the requested personal data. Solove (2013) shows in his analysis that "the basic approach to protecting privacy has remained largely unchanged since the 1970s" (1880) and that the consent approach to privacy has ceased to be meaningful in the contemporary networked information society. The basic assumption behind consent is that "data subjects make conscious, rational and autonomous choices about the processing of their personal data" (Schermer, Custers, and van der Hof 2014, 171), which is under pressure in today's information and data-driven business practices. In these practices, people consent to providing personal information without much thought, without having read or fully understood the consent form, and, furthermore, consent forms are often rather lengthy and written in legalistic language that few people understand.

The datafication of personal information constitutes a new kind of information society. Digitization was the process of taking the analog world to the digital environment; it allowed society to store more information and process it more rapidly. In the digitized era, digital information was still treated as if it was analog, and it was often used within the same "singular purposes" (Mayer-Schönberger and Cukier 2013, 15), for which it was collected and to which its "value was tied" (15). The next step is to datafy information, to "put it in a quantified format so it can be tabulated and analyzed" (78). Datafication allows analysis of information in more sophisticated ways and allows analyses across large data sets. It breaks down the traditional understanding of data as numbers and information as texts, movies, music, and so on. A good example is Google Books. When Google digitized books, it scanned them in a way that allowed for full-text searching and stored the text in a way that allowed people to search for particular words or phrases across millions of books in a few seconds. Such searches can compare the appearances of certain words and phrases in books from early publications up to publications today (https://books.google.com/ngrams). Google has gone beyond the mere digitization of the books; it has datafied them—enabled analyses of the content of books as though the content is data. In this sense, the traditional data–information–knowledge–wisdom (DIKW) pyramid breaks down; in the DIKW tradition, data are viewed as singular symbols without meaning, whereas information is processed data that have gained meaning (cf., e.g., Bellinger, Castro, and Mills 2004). When something is datafied there is no distinction between what data are and what information is: These are all elements that can be analyzed for patterns and correlations. Purchases at Target as well as books scanned by Google can be datafied—and everything in between.

Given that most digital devices are connected to the Internet today, and more will become connected in the near future, and given that many of our daily activities are digitally mediated and connected, datafication might in the future include "everything," which "is not as far out as it sounds" (Mayer-Schönberger and Cukier 2013, 94). Once everything and all activities are digital "the potential uses of the information are basically limited only by one's ingenuity" (96). These vast possibilities for analysis are currently driving many businesses "where big data is being used to create new forms of value" (97). As more organizations realize the potential in collecting, datafying, and analyzing information created by users' and customers' interactions with digital media, datafication and predictive analysis will become more predominant (Kerschberg 2014). The more personal information organizations have about their users and customers, the better services and more precise advertisement they will be able to offer. Shoe stores, gas stations, the social benefits office, intelligence services, tax authorities, search engines, universities, hospitals, and social networking sites all share the same common goal of providing services, and the quality of their services is dependent upon the amount of knowledge they have about their users and customers. The more personal information they have, the better the

service. If they were able to share information among each other, their service could be even better. If the shoe store had information about my tastes in fashion, which I share on my social networking site, the shoe store would be able to provide even better service to me. If the social benefits office had information about my health history, it would be able to provide better service. And so on.

The basic premise of this line of thinking ought to be questioned—more information is better, more digital services improve people lives, and that greater connectivity provides a better world. As Winner (1986) noted, our thinking is dominated by "mythinformation: the almost religious conviction that a widespread adoption of computers and communications systems along with easy access to electronic information will automatically produce a better world for human living" (105). Winner called for caution in the development of the information society. While the benefits of modern information and communication technology might be obvious to many people, it is important that the development of technology and information policies takes fundamental human rights and societal concerns into account. One challenge to this kind information society is the type of privacy protection people are granted in this brave new world of datafication of personal information. There is little doubt that the creation of big data businesses challenges privacy; the question is whether the problem merely changes in terms of scale—that privacy is more at risk in the big data society—in which case we merely need to increase the existing efforts and privacy protections. However, "if the problem changes, we may need new solutions" (Mayer-Schönberger and Cukier 2013, 153); in other words, if the big data society fundamentally changes our approach to data and dealings with personal information, then we need to ascertain what privacy means in the era of big data and datafication.

The contemporary ethical challenge in the big data age is not whether to collect personal information. The fact is that personal information is being collected and stored by private corporations and public agencies as we interact in the digital environment. The challenge is to determine when and how it is ethically responsible to analyze the information, what to look for in the data, which questions to ask of the data, and the scale to which it is reasonable to make predictions about future events and actions based on that data.

Privacy: Kinds and challenges

Since Warren and Brandeis ([1890] 2005) formulated the basic principle of the right to be left alone, the notion of privacy has been reconceptualized a number of times, often due to advances in technologies and developments in the commercial market. Tavani (2008) suggests that there are four different kinds of privacy, each of which has its own unique focus. Warren and Brandeis's ([1890] 2005) notion of privacy can be called "physical privacy" (Tavani 2008, 135), which is "freedom from (physical) intrusion" (135). The second and third, respectively, are "decisional privacy," which is "freedom from interference affecting" (136) important decisions, and "psychological privacy," which is concerned with the protection of "one's intimate thoughts" (137). Further discussion and analysis of these kinds of privacy are beyond the scope of this article, though I note that the distinctions are not as clear as they are sometimes presented to be.

The fourth one, informational privacy, is of special interest in this article. This kind of privacy is often associated with Westin's (1967) definition from the 1960s: "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (7). As mentioned earlier, two common approaches to conceptualizing informational privacy are "restricted access" (Tavani 2008, 141) and the "control theory" (142). In the "restricted access" tradition, an individual "is able to limit or restrict others from information about" him- or herself (141). The basic idea here is that an individual is able to set up private zones or contexts in which personal information is held and the individual may enjoy privacy when he or she is able to restrict or limit others from accessing personal information held in those zones or contexts. The "control theory" tradition is somewhat related to the "restricted access" tradition; the control theory tradition is concerned with an individual's ability to control who has access to information about the individual. A central element in the control tradition is "the role that individual choice plays" (143) and the ability of individuals to control whether access to information about themselves is granted or restricted to others.

A number of critiques have been introduced that problematize these conceptions of informational privacy. I touch on five of them to highlight some of the conceptual challenges associated with the notion of informational privacy. It is beyond the scope of this article to address the challenges in detail and propose an alternative formulation of the notion of informational privacy. The purpose of the present discussion of these five challenges is to help us conceptualize the notion of datafication of personal information. While some of the challenges remain with the proposed datafication model of personal information, a crystallization of the challenges lays the conceptual foundation for the proposed model. I return to that later in this article.

The first challenge is that the notion of "information" in informational privacy is unclear. While many authors

refer to the notion of information, few make it explicitly clear what they are indeed talking about. Solove (2008) cites Murphy's definition that personal information is "any data about an individual that is identifiable to that individual" (25), which is either a rather broad or a rather narrow definition. It is a broad definition in the sense that there is much information about us that we might not necessarily deem private: my profession, for instance, or any information that may be gathered by merely looking at me when I walk down the street. On the other hand, it could also be regarded as a narrow definition if personal information is only that information that uniquely identifies a specific person—then only information that is specific to me is my personal information, my Social Security number, for instance, and perhaps my name and address.

Floridi (2005) takes this idea one step further and suggests that:

> "My" in "my information" is not the same "my" as in "my car" but rather the same "my" as in "my body" or "my feeling"; it expresses a sense of constitutive *belonging*, not of external *ownership*, a sense in which my body, my feeling and my information are part of me but are my (legal) possessions. (195, italics in original)

In Floridi's sense of personal information, people do not own their personal information, they are their personal information. Personal information is not something that can be owned. Personal information is therefore not external information about a person—it is that person. My home address is therefore not personal information ("my" address is temporary, like "my" car), and neither is information about my height, profession, hair color, marital status, income, taxes, and so on. Only a very limited set of information is "my" information in Floridi's sense of the concept. Solove (2008), however, takes a different and more pragmatic approach when he reminds us that "personal information is often formed in relationships with others. All parties to that relationship have some claim to the information" (27). When I purchase gas, that information belongs both to the gas station and to me. It is information about me in the sense that it says something about my choices (how much gas, how often, which brand, etc.) and it is important information for the gas station (when did it sell the gas, how much, how was it paid for, etc.).

When discussing how and when to limit access to or to control information, it makes a difference whether we are talking about Murphy's information as identifying a specific person, Floridi's sense of "my" information, or Solove's pragmatic notion of information as relational. To leave the definition open or up to common sense runs the risk of creating too much unnecessary confusion in the conversation about informational privacy.

Second, informational privacy's focus is on individuals' abilities to limit access to or control their personal information and is as such less focused on the "social question" (Lyon 2001, 150) that surveillance societies create. Informational privacy is concerned with setting private zones and determining how individuals control information about themselves. Surveillance, however, is "a means of sorting and classifying populations and not just of invading personal space or violating the privacy of individuals" (Lyon 2001, 151). In fact, the history of the modern information society is built on the struggle to limit the state's (and commercial corporations') ability to surveil citizens and protect their liberties, as noted by Westin (1967): "Surveillance is obviously a fundamental means of social control. … One of the central elements of the history of liberty in Western societies since the days of the Greek city-state has been the struggle to install limits on the power of economic, political, and religious authorities to place individuals and private groups under surveillance against their will" (Westin 1967, 57).

With the focus on singular pieces of information and the rights of individuals to control or restrict access to that information, the privacy literature runs the risk of losing sight of the larger societal power relations. With the explicit incorporation of the purposes of surveillance into the notion of privacy, the focus might broaden to include sorting and classification as core elements in privacy theory. In this sense, privacy is not solely about individual pieces of personal information that the individual wishes to control; privacy concerns also arise as the information is produced about individuals as they are sorted and classified for specific purposes.

The third challenge is closely related to the second: Informational privacy is often conceptualized as the ability of individuals to protect themselves, though it could be argued that there is something more fundamental to privacy that reaches beyond the individual. Hosein (2006) suggests that the choices an individual makes to protect his or her personal information are situated in a greater conceptual framework of thinking grounded in human dignity:

> The point to understanding privacy as a core feeling, something inexplicable, is that it is tightly intertwined with our sense of right and wrong, our moralities. Even beyond morality, beyond the debates of relativism and absolutism of norms and morals, there is something about human dignity. (124)

In this line of thinking, privacy is a fundamental human right. The right goes beyond specific legal frameworks, technologies, and contexts. The individual-centered approach to informational privacy emphasizes the individual's responsibilities and focuses on the individual's ability to control

their information. A human rights approach suggests that the right to privacy is beyond the individual and not something that the individual should have to work for. Frohmann (2008) asks a more fundamental question of the ethical foundation of informational privacy, namely, whether technologies and practices go beyond individual choice and "bypass the epistemological subject because they bypass consciousness" (270). If informational privacy is conceptualized purely as individual choice it misses out on a great number of significant moves in contemporary society that go beyond the individual's ability to make conscious decisions about their own informational-self. The focus on the rational individuals to limit or control information about themselves is not only impractical—given the large amount of information that most people in the Western world handle on a daily basis—it also misses out on the power structures at play in contemporary networked information society: Much of the information about individuals is not provided by anyone, but is produced through predictive analytics. The third challenge is therefore to consider what informational privacy would look like if the focus were not on the individual's ability to limit or control personal information but on human dignity or power structures beyond the reach of individuals.

The fourth challenge has received a great deal attention in privacy literature, but has had a constraining effect on the conceptualization of informational privacy, namely the distinction between private and public. As noted by DeCew (1997), the public–private dichotomy has sometimes been used to make a distinction between "the appropriate scope of government as opposed to self-regulation by individuals" (10) and at other times to "differentiate political and domestic spheres of life" (10). The general assumption has been that there is a "boundary marking off that which is private from that which is public" (10). While the distinction between public and private does not in itself define the notion of privacy, it has generally framed the common understanding of privacy in the sense that people ought to enjoy privacy when in private, whereas they may not enjoy the same protection when in public. Following the same thinking, it is assumed that information is either private or public depending on the sphere in which it is located, and that all information not located inside people's private spheres is public information. The dichotomy can cause confusion. Zimmer (2010), for instance, shows how some researchers who gathered information for their research project on Facebook assumed that the information "was already publicly available" (321), and Nissenbaum (2010) suggests that:

> the dichotomy of information is surely so deeply entrenched in the public's thinking about privacy that it will resist even the seismic shifts in information systems and practices caused by digital information technologies and media. (119)

While the public—private dichotomy might help to conceptualize physical privacy—to determine the boundaries of one's home—the dichotomy adds more confusion than help when it comes to informational privacy. In contemporary networked digital information society, people sit in their private homes, connected to a public network, communicating with private friends, using public wires, exchanging private information, stored on public servers. In such a society, the distinction between public and private surely becomes blurred.

Lastly, the fifth challenge is that the very value of privacy has not been clearly articulated in a systematic and general way that has reached consensus. Just as the definition of privacy is open for interpretation and discussion, the value of privacy is open for interpretation and discussion. Floridi (2005) offers two "popular" (193) theories that explain the value and importance of privacy: "the *reductionist interpretation* and the *ownership-based interpretation*" (193, italics in original). Both theories are, according to Floridi, conceptually flawed.

The reductionist interpretation views privacy as a utility to protect other interests or as a protection against undesirable consequences. These interests and consequences, however, are not always articulated and stated in a clear and objective fashion. Floridi (2005) questions the assumption that more privacy necessarily leads to better and healthier societies and asks whether "the defence of informational privacy in the home may actually be used as a subterfuge to hide the dark side of privacy: domestic abuse, neglect or mistreatment" (194). In other words, for a reductionist interpretation of privacy to be successful, it needs a clear articulation of the consequences that privacy shall protect or value. In this approach, privacy is not a value in itself—it is a means toward other values.

In the ownership-based interpretation of privacy, personal information is viewed as something that people own and that "each person's right to bodily security and property" (Floridi 2005, 193) shall be respected. In other words, given that a person owns his or her personal information, that person has a right to the privacy of that information. The main challenge to this notion of privacy is that much information is relational in the sense that more than one person owns that information. Both the gas station and I own information about my purchases at the gas station. The value and special nature of informational privacy is not solely explained by the status of the ownership; the ownership of information might have little to do with the protection of privacy.

While there are many challenges to the current conceptualization of informational privacy—Solove (2008, chap. 2, 12–38) and Nissenbaum (2010, Part II, 67–126) review many of them—these five challenges are especially pertinent when it comes to the datafication of personal information. When predictive analytics is applied in the big data cloud of personal information, these five challenges remain—but they are reconfigured. I highlight this reconfiguration of the five challenges next:

1. The information generated by predictive analytics in big data sets is new information. The challenge, therefore, is not merely to restrict or control access to personal information, because the production of the new information about oneself is beyond one's control. When personal information is datafied and new personal information is produced, it is not clear within the current informational privacy frameworks who owns this information or has the right to it.
2. One goal of big data analysis is to classify and sort people. Commercial enterprises and state agencies have interests not necessarily in specific individuals, but often in large groups of people and the characteristics that these groups exhibit. The protection against being classified and sorted for organizations' purposes and goals ought to be a focus for informational privacy—such classification and sorting goes beyond the control and restriction of access to individual pieces of personal information, but is instead about the characteristics and clustering of personal traits and habits.
3. The analysis and production of new information about individuals is beyond their control in the big data regime. An individual does not have the power or intellectual capacity to control the flow of information into and among information empires and data-brokers, the generation of big data, and the knowledge produced by statistical computation.
4. It is possible to make a separation between private spheres and public spheres in the analog world; it becomes more difficult in the digital world, and nearly impossible in the networked information society. As people use computers and network technologies to perform many daily activities, the distinction between what is done in a private sphere and what is done in a public sphere becomes almost impossible to distinguish, both from a technological viewpoint and from a human perspective.
5. Regardless of recent technological advances, be it the "instantaneous photographs" that Warren and Brandeis ([1890] 2005, 210) were concerned with or the predictive analysis in today's big data cloud, it appears that privacy is valued at the foundational level where it is, "tightly intertwined with our sense of right and wrong, our moralities" (Hosein 2006, 124). It may be that a "society devoid of any informational privacy" (Floridi 2005, 193) may be a society with no crime, fraud, or missing people, but it will also be a society with no freedom, as Orwell (1949) so convincingly showed us.

Privacy models

In the age of big data we need to think differently about privacy. We need to shift our thinking from definitions of privacy (characteristics of privacy) to models of privacy (how privacy works). Moreover, in addition to the existing models of privacy—the surveillance model and capture model—we need to also consider a new model: the datafication model presented in this article.

Agre (1994) calls for shift in our conceptualization of privacy from a "surveillance model" to a "capture model." Both of these models operate with their distinct metaphoric components, which shape how we discuss and describe the phenomenon.

The traditional "surveillance model" has five components:

1. Visual metaphors, as in Orwell's "Big Brother is watching you" or Bentham's Panopticon.
2. The assumption that this "watching" is nondisruptive and surreptitious.
3. Territorial metaphors, as in the "invasion" of a "private" space, prototypically the family home, marked out by "rights" and the opposition between "coercion" and "consent."
4. Centralized orchestration by means of a bureaucracy with a unified set of "files."
5. Identification with the state, and in particular with consciously planned-out malevolent aims of a specifically political nature (Agre 1994, 105–106).

Observing the developments in the computer technology, Agre (1994) called for a different model of privacy—the "capture model" (Agre 1994, 107). This model's name itself indicates two interrelated notions at play: (i) the epistemological notion of capturing information and data, and (ii) an ontological notion of modeling the reality that the information or data reflect. Agre thus contrasted the capture model with the surveillance model:

1. Linguistic metaphors for human activities, assimilating them to the constructs of a computer system's representation languages.
2. The assumption that the linguistic "parsing" of human activities involves active intervention in and reorganization of those activities.

3. Structural metaphors: the captured activity is figuratively assembled from a "catalog" of parts provided as part of its institutional setting.
4. Decentralized and heterogeneous organization: the process is normally conducted within particular, local practices that involve people in the workings of larger social formations.
5. Driving aims that are not political but philosophical, as activity is reconstructed through assimilation to a transcendent ("virtual") order of mathematical formalism (Agre 1994, 107).

It is beyond the scope of this article to describe and discuss the details of Agre's capture model of privacy, although it should be noted that a key notion of the capture model is that the capturing of data is never neutral or "purely technical but always sociotechnical in nature" (Agre 1994, 112).

The traditional surveillance model of privacy assumes that the passive watching and acquisition of data will be not be affected by the collection of the data. It therefore presumes that the collected data are a mere reflection or rational representation of reality. The capture model takes a different stand; it assumes that technologies change or affect social activities and that the simple capturing of data does modify the way people interact with technology. It therefore presumes that that the captured data are an oversimplification of reality. However, both models share a common focus on the collection of data and both are concerned with how and for what purposes the collected data are later used.

The proposed datafication model of privacy shifts the focus from data collection to data processing and analysis. Data collection is ontologically oriented; it focuses on data as representing facts about states of affairs in the world: people and activities and the interrelation between places, times, other people, activities, and intentions. Data processing and analysis is epistemologically oriented; it focuses on the facts or realties that data can generate as they are processed and analyzed.

The surveillance model and the capture model are both at their core concerned with privacy protection in the data collection tradition as a means of control of data gathered through surveillance activities or capture activities. The datafication model, on the other hand, assumes that data have been collected, amassed, stolen, bought, hacked, or otherwise acquired and that the privacy concerns at play are the construction of new knowledge, insights, or realities based on the available data. The woman who had scored high on Target's pregnancy prediction score might have felt that her privacy had been violated, but neither the surveillance model nor the capture model can fully explain that violation because both are focused on the collection of data. The woman's privacy was not violated due to the collection of data (she had presumably volunteered the information), but it could be argued that her privacy was violated due to data processing and analysis. This is what the datafication model of privacy aims to achieve.

By shifting focus from data collection to data processing and analysis, the privacy concern is reconfigured from which facts entities know about people to which facts the entities have produced about people. Where the surveillance model of privacy achieved its main characteristics via the metaphor of watching, and the capture model by the sociotechnical notion of grammar of action, the datafication model of privacy can be characterized by the metaphor of patterns of behavior.

The three models of privacy—surveillance, capture, and datafication—supplement each other; they are not competing understandings of privacy. They are different models that help us to understand and appreciate the notion of privacy—each model takes a different viewpoint and focuses on specific features of the sociotechnical phenomena that are under investigation. As such, the surveillance model focuses on the tension between the watchers and the watched, between the public and private spheres, and on the present power relations. The capture model focuses on the codification of activities, the sociotechnical nature of computer technology, and on the unclear purposes of data collection. The datafication model focuses on the anonymous creation of new personal information, the reinterpretation and statistical analysis of data, and the commoditized nature of personal information. Taken together, the three models provide a powerful holistic approach to privacy in the networked information society; individually, they stress different aspects and highlight different features.

The datafication model does not address the five challenges to privacy but directly addresses or speaks to them. Let me briefly return to the five challenges:

1. The kind of information that the datafication model addresses is new information produced by data processing and analysis. The information may or may not be true, and the information is in all situations relational information about identifiable individuals or entities.
2. The datafication model is concerned with the ability to classify and sort people based on the available data—and thereby to create new insights and correlations between people, their activities, and interests. The information in question is beyond individuals' control or ability to limit access to the information, because the information is not—and has never been—in their possession.

3. While it is beyond the abilities of individuals to control and limit the personal information that they leave in the digital marketplace, the datafication model shows that this information is being processed and analyzed and that it falls within the interest of privacy.
4. By shifting focus from where and how information is collected or produced (in the public or private spheres) to the information itself, the datafication model suggests that there are privacy interests at play where organizations use data to produce new personal information that people have not volunteered to the organizations.
5. As people interact online, they consequently leave behind personal information, which might be a fair price to pay at the time and which might leave some to proclaim the end of privacy. The datafication model suggests that there are privacy concerns at play when organizations process and analyze the digital footprints left behind.

The datafication model of privacy supplements Agre's surveillance and capture model of privacy and allows for analysis of the privacy of personal information in the age of big data. Privacy concerns in the age of big data also arise out of organizations' ability to construct new personal information through predictive analytics. The datafication model of privacy highlights this feature and provides a framework for analyzing privacy concerns in big data.

Conclusion: Datafied personal information

The age of big data calls for a reconceptualization of the notion of privacy. Previous models of privacy limit their focus on collection and gathering of data as the central mechanism of the privacy concern. Accordingly, privacy is seen as the ability to restrict access to information or the ability to control the flow of personal information. In the age of big data, a significant concern is how new personal information is produced by businesses and organizations through predictive analytics.

This article presented a new conceptualization of privacy, the datafication model of privacy, which shifts focus from data collection to data processing and analysis. Herein focus is thereby shifted from concerns about revealing information about oneself to others to concerns about the new insights that others can generate based on the already available data.

References

Agre, P. E. 1994. Surveillance and capture: Two models of privacy. *Information Society* 10 (2):101–127.

Allen, A. L. 2013. An ethical duty to protect one's own information privacy? *Alabama Law Review* 64 (4):845–866.

Bellinger, G., D. Castro and A. Mills. 2004. Data, information, knowledge, and wisdom. http://www.systems-thinking.org/dikw/dikw.htm (accessed September 12, 2015).

DeCew, J. W. 1997. *In pursuit of privacy: Law, ethics, and the rise of technology*. Ithaca, NY: Cornell University Press.

Duhigg, C. 2012. How companies learn your secrets. *New York Times*, February 16. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (accessed September 12, 2015).

Floridi, L. 2005. The ontological interpretation of informational privacy. *Ethics and Information Technology* 7 (4):185–200.

Frohmann, B. 2008. Subjectivity and information ethics. *Journal of the American Society for Information Science and Technology* 59 (2):267–277.

Hosein, G. 2006. Privacy as freedom. In *Human rights in the global information society*, ed. R. Frank Jørgensen, 121–147. Cambridge, MA: MIT Press.

Kerschberg, B. 2014. Five steps to master big data and predictive analytics in 2014. http://www.forbes.com/sites/benkerschberg/2014/01/03/five-steps-to-master-big-data-and-predictive-analytics-in-2014/ (accessed September 12, 2015).

Lyon, D. 2001. *Surveillance society: Monitoring everyday life*. Buckingham, UK: Open University Press.

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

Nissenbaum, H. 2010. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

Orwell, G. 1949. *Nineteen eighty-four*. London, UK: Penguin Books.

Schermer, B. W., B. Custers and S. van der Hof. 2014. The crisis of consent: How stronger legal protection may lead to weaker consent in data protection. *Ethics and Information Technology* 16 (2):171–182.

Solove, D. J. 2008. *Understanding privacy*. Cambridge, MA: Harvard University Press.

Solove, D. J. 2013. Privacy self-management and the consent dilemma. *Harvard Law Review* 126:1880–1903.

Tavani, H. T. 2008. Informational privacy: Concepts, theories, and controversies. In *The handbook of information and computer ethics*, ed. K. E. Himma and H. T. Tavani, 131–164. Hoboken, NJ: Wiley.

Warren, S. D. and L. D. Brandeis. (1890) 2005. The right to privacy. In *Information ethics: Privacy, property, and power*, ed. A.D. Moore, 209–225. Seattle, WA: University of Washington Press.

Westin, A. F. 1967. *Privacy and freedom*. New York, NY: Atheneum.

Winner, L. 1986. *The whale and the reactor: A search for limits in an age of high technology*. Chicago, IL: University of Chicago Press.

Zimmer, M. 2010. "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology* 12 (4):313–325.