Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 71-89

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

Open Traffic Data for Future Service Innovation – Addressing the Privacy Challenges of Driving Data

Anna Rohunen,[1] Jouni Markkula,[2] Marikka Heikkilä,[3] and Jukka Heikkilä[4]

[1] University of Oulu, Department of Information Processing Science, Oulu, Finland, anna.rohunen@oulu.fi,
[2] University of Oulu, Department of Information Processing Science, Oulu, Finland, jouni.markkula@oulu.fi
[3] Turku School of Economics, Centre for Collaborative Research, Turku, Finland, marikka.heikkila@utu.fi,
[4] Turku School of Economics, Centre for Collaborative Research, Turku, Finland, jups@utu.fi

Received 1 August 2013; received in revised form 2 February 2014; accepted 6 March 2014

Abstract

Following the present open data policies, traffic data are collected and increasingly made openly available by different organizations. Yet, expanding use of mobile technologies with tracking possibilities provides means to collect precise and rich information about individual vehicles and persons in traffic. This personal driving data, combined with other open traffic data, have a great potential for future open service innovation. However, information privacy presents a major challenge for collection and efficient utilization of the data. In this paper, we present a view of the near future development of personal driving data collection and usage for open traffic data production by addressing the privacy challenges. We review the existing privacy behavior models and present our empirical findings from driving data based service pilot studies. Our results show that, despite their privacy concerns, the data subjects are willing to disclose driving data for services, especially for some benefits in return. We identified the following key factors affecting data disclosure: informing of personal data processing, trust in organizations of the service ecosystem, and users' control over their data. Understanding of these factors helps mitigating the users' privacy concerns when personal data based services are designed and production of open data is planned.

Keywords: Open data, Service innovation, Data privacy, Privacy concerns, Privacy behavior models

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 71-89

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

1 Introduction

Our economies are moving globally to the stage of a knowledge economy, where value is generated by information for societies, ecosystems, organizations, and users [10]. In the knowledge economy, development is driven by creativity and open tools. New innovations are produced by following an open innovation paradigm [6], [7] in business ecosystems [39]. Governments understand the value of open data for the development of the information society. They are promoting open-data policy and enforcing the re-use of public sector information via legislation and other administrative means. In the European Union, this is explained by the Public sector information (PSI) Directive [14], [18], and in the United States, by the Open Government Directive [51]. An essential part of public sector information is spatial information, and its more efficient usage is supported through specific legislation—in Europe, with the Directive on Infrastructure for Spatial Information in the European Community (INSPIRE Directive) [16]. Some widely used and open public sector spatial data resources are national road and street databases, such as Digiroad (Site 1) in Finland, which are used as a basis for various traffic information services provided by both public and private sector organizations.

New advancing information and communication technology (ICT) has provided continuously better technological means to collect, process, and utilize data in various forms. For example, current Internet and mobile technologies allow the collection of real-time, location-based data from an increasing number of various sensor networks, vehicles with on-board global positioning system (GPS) devices, and even from private persons with positioning-enabled mobile phones. Voluminous data can also be processed more efficiently to produce and deliver information services to users in real time. Data produced by tracking of vehicles can be utilized by governments and other organizations responsible for traffic planning, as well as to provide information services to individual people in traffic. On the other hand, as the number of potential usages of vehicle tracking data increases and the number of parties interested in using and benefitting from the secondary usage of data rises, privacy concerns are becoming more serious.

An example of constant evolution of technical means and novel uses of data is the case of German road tolls (see, e.g., [40]). Camera toll booths came in use on the Autobahn in 2005; later, on-board devices with GPS and Internet connections were added. The functionality expanded from toll booths to re-routing, pay-by-drive, payment authorization, identifying and locating vehicles and fleets, detecting cargo and toll evasion, and so on. In other words, the data originally gathered from the vehicles for the purpose of road tolls ended up serving purposes that were unforeseen when the system was launched. The privacy issue was also made explicit in a recent discussion in Finland: In December 2013, a working group from the Ministry of Transport and Communications proposed satellite monitoring–based kilometer taxation on car use in Finland [38]. The group stated that before the final decision to adopt the new national monitoring system, however, privacy protection needed to be ensured. The report [37] also stated that for any other usage of vehicle tracking, such as pay-as-you-drive insurance services, explicit informed consent from the data subjects should be required.

There is an evident need for academic discussion on privacy issues related to driving data. Open data policies and their potential incongruity with privacy and data protection legislation have also been brought up recently in scientific discussion (see, e.g., [31]). However, empirical knowledge about this issue from the point of view of data subjects is still limited, especially since there has not been much real-world experience of actual driving data collection and privacy concerns associated with it. Correspondingly, although the literature studying attitudes and behavioral intentions of users to disclose their personal data is abundant (e.g., [8], [34], [35]), literature analyzing willingness to disclose personal data in real situations (e.g. [29], [45]) is scarce. Furthermore, the previous studies have analyzed individuals' privacy related behavior in the context of different Internet and e-commerce services, and presented various privacy behavior models based on that (see, e.g., [33]). However, studies of privacy behavior from the point of view of producing open data to generate new service innovations are scarce.

The objective of this paper is to present the current view of the near future development of driving data collection and usage for governmental purposes, such as taxation, and its potential secondary usage for open traffic data, especially from the perspectives of data privacy and data subjects' point of view. To accomplish this, we first discuss driving data collection and its usage for producing open traffic data. Next, we present a review of privacy behavior models summarizing the current literature on individuals' willingness to disclose personal information and their privacy concerns. Third, we provide the results of our empirical studies on data subjects' privacy concerns and their willingness to disclose personal information in the context of personal driving data. Our findings on the conditions in which drivers are willing to disclose their personal driving data and relevant factors affecting their privacy concerns can be helpful when designing the collection of driving data to produce open traffic data.

The remainder of this paper is organized as follows. The next section presents the application context of the research, defining and describing personal driving data and their processing for open traffic data. Following this, a review and analysis of privacy behavior models on willingness to disclose private data is presented in section 3. In section 4, our empirical research and findings on driving data and privacy concerns are presented. Next, in section 5, the application of our results for improving the practices of collecting personal driving data with consent from the data subjects is discussed. Finally, section 6 concludes the paper.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 71-89

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

2 Future Open Traffic Data and Privacy Challenges

The current trend of opening public sector data, supported by governments' policies and facilitated by the advancement of technology in data-intensive services, has resulted in the increasing release of open data–based traffic-related services. Many cities are publishing real-time or near-real-time traffic information, such as in the New York City (Site 7) and Oulu city (Site 10) traffic maps. Map data and services with navigation support are available, for example, from Google (Site 5) and OpenStreetMap (Site 9). Some forms of traffic data produced by the public authorities are also openly available, such as Digitraffic (Site 2) in Finland. Available spatial data, including map and traffic data, are collected and offered to the user by public sector organizations and companies, as well as by third-sector non-governmental organizations (NGOs). Public sector traffic data is mostly collected with various sensors that produce data for traffic management systems; in addition to their original main purpose, they can also be used for other reasons. When data are opened, other parties can use them to innovate completely new services that support more intelligent traffic and transport systems.

While widely available open traffic data currently exist, such information is largely based on infrastructural sensors and often collected with rather old technology, for example, vehicle detection with magnetic loops installed in the roads. However, future intelligent transport systems (ITS) are seen to be essentially based on the real-time tracking of vehicles in traffic with positioning technologies and wireless communication. This allows both more advanced management of traffic systems by the public sector and improved services for the individual in traffic. When driving data based on tracking of vehicles start to be collected for the purposes of public authorities, this practice falls under the same open data policies as other public data. By producing open driving data and making them widely available, genuine possibilities for more advanced intelligent traffic services developed by various innovative organizations are generated. Driving data allow for novel types of services that are not possible based on any other form of old traffic data.

The foundations for the future directions of vehicle tracking–based driving data are laid, for example, in the European Union ITS Directive [17] and the European Electronic Toll Service (EETS) Directive [15]. Future solutions are projected to involve positioning that will enable on-board devices in vehicles that produce data for traffic management systems, as well as for road tolls and other pay-as-you-drive types of services such as insurance. Such devices have been proposed as the means to diminish pollution (as a better alternative to fuel taxation) or congestion, noise, and unnecessary driving in general (as a better alternative to road tolls, or their electronic equivalent tollbooths, such as EZPass (Site 3), or FasTrak (Site 4). For example, a number of trials of road tolls have been carried out around the world. The results show that the technology for tracking individual vehicles does work, and significant advantages can be achieved by introducing more precise road pricing tolls [53]. It has been stated that a "clear system of incentives is critical to changing driving behavior" [41]. This is also indicated in the case of Stockholm road tolls, where a minor road toll had a huge impact on downtown traffic congestion (see, e.g., [19]). Decisions to introduce taxation mechanisms based on tracking, however, are very political. German trials ended up with the introduction of road tolls for heavy transports only (LKW-MAUT; see, e.g., [40]), because controlling heavy traffic was generally accepted (noise, emissions, high load on roads, congestion, and tax/toll evasion). In the Netherlands, the initiative was stopped by the new government.

In addition to public sector applications, there are already commercial services available based on driving data. For example, OnStar (site 8) in North America allows some insurance discount solutions based on the collected driving data. A corresponding, but functionally quite different, system is provided by Helpten (Site 6), which offers, for example, driver's logs and driving habit analyses for its customers. This type of system can be called a driving data multiservice system (DMS), as it provides several different services. In DMS systems, driving data are collected, managed, and processed by a driving data operator. Personal driving data can then be disclosed to the drivers themselves, or to some other parties in the business ecosystem that can offer services to customers (e.g., road toll services or insurance companies). The collected data also have the potential to be used for producing open traffic data, if the driving data operator fulfils the legitimate and fair personal data collection and processing requirements [36].

One of the biggest challenges limiting the evident utility of driving data for producing open traffic data–based service innovation is information privacy. As noted by Kulk and Loenen [31], for example, open data policies are potentially in conflict with information privacy; this is very much the case when it comes to driving data. Driving data reveal information about the individual persons' behavior, which clearly falls under the data protection legislation. However, this obstacle may be overcome if adequate privacy protection methods and practices are followed and further developed.

Information privacy is protected in many countries by legislation; for example, in the European Union, the European Commission Data Protection Directive [13] and corresponding national legislations are in force. Privacy laws are generally based on widely accepted international privacy guidelines and principles, for example, the Organisation for Economic Co-operation and Development (OECD) Guidelines on Privacy [42] and the United States Federal Trade Commission (FTC) Fair Information Practice Principles [20]. All of the following are similar general principles that are supported widely at an international level: Personal data must be collected only for specified, explicit and legitimate

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

purposes; the data may be processed only if the data subject has unambiguously given his/her consent or processing is necessary (e.g., for the performance of a contract to which the data subject is party or for compliance with a legal obligation to which the controller is subject); the data must be processed fairly and lawfully; the data controller and processor need to apply confidentiality and security in the processing of the data; and they need to implement appropriate measures to protect personal data against accidental changes and unauthorized disclosure and access.

Even if the general privacy principles have stayed quite similar for a half century, the new technologies for data collection and utilization have changed the conditions for applying them. Therefore, a need to update the legislation has emerged. This can be seen, for example, from the currently ongoing revision reform of European Union data protection directive [12] and related discussion. Furthermore, our practical understanding of different aspects of information privacy as a part of information systems design has improved. The privacy by design (PbD) concept has contributed significantly to this understanding; this has been developed and promoted by Information and Privacy Commissioner of Ontario, Canada, Ann Cavoukian [2]. This is a concept that demands a proactive approach to privacy by embedding it in the system from the initial stage until the end of the system's lifecycle [3]. PbD is founded on the following seven principles:

1. Proactive not reactive; preventative not remedial
2. Privacy as the default setting
3. Privacy embedded in the design
4. Full functionality—positive-sum, not zero-sum
5. End-to-end security—full lifecycle protection
6. Visibility and transparency—keep it open
7. Respect for user privacy—keep it user-centric.

Furthermore, the PbD concept extends privacy solutions to the following three areas, which Cavoukian calls the *Privacy by Design Trilogy*:

- Information technology
- Accountable business practices
- A physical design and networked infrastructure

Over the past few years, research has paid most attention to the information technology aspect of privacy solutions. This essentially includes the development of Privacy-Enhancing Technologies (PETs). An overview of the current state of PET research and practice can be reviewed, for example, from the latest proceedings of the Privacy Enhancing Technologies Symposium (PETS) [9], [19], [21], [22] and the Center for Internet and Society's PET Wiki [4]. Corresponding, and highly relevant, knowledge on data disclosure control methods is also available in the field of statistics [54], [55]. However, in developing the PbD concept, it has been noted that PETs, which are limited to information technological solutions, are not sufficient for solving the multifaceted problems of information privacy.

Considering the future direction of driving data collection to produce open traffic data, the understanding and development of accountable business practices are vital. The computing infrastructure gets more complex generating more precise, automatically collected information, and pressures to open data for unforeseen purposes grow in the name of environmental protection and safe and more efficient traffic. When valuable driving data collected for specific purposes, for example, taxation or road tolls, is used for secondary reasons such to produce open data, explicit consent is needed from the data subjects. This calls for accountable business practices, and the consent should be *informed*; that is, the data subjects should understand what information is collected, how it is processed, and where it is disclosed for what purposes. This implies transparency of data processing on the part of the data collectors and all the parties involved in the related business ecosystem. Ideally, this transparency should allow the whole process of personal data management to be clarified in such a way that responds adequately to the data subjects' privacy concerns. To achieve this, we need to understand what their actual privacy concerns are and in what conditions they are willing to agree to disclose their personal data. Next we shall first give an overview to privacy behavior models in general and thereafter reflect the traffic environment in particular based on our empirical studies on the subject.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 71-89

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

3 Literature Review on Privacy Behavior Models

Information privacy and related personal behavior have emerged as an active research topic as the Internet has developed. These concerns have become increasingly important due to the advancement of technology, as well as the availability of new types of personal data and applications. A number of studies have been published on individuals' privacy behavior, specifically their intentions and actions in using Internet services, electronic commerce, or other services where personal data is collected. To the best of our knowledge, privacy behavior theory has not previously been applied in the context of open traffic data [56].

The application context in which these models have been developed or studied typically involves specific commercial Internet–based services where personal data are potentially collected. For example, in their research, Liu et al. [34] studied behavioral intention for online transactions and how they are affected by privacy and trust, which typically means "faith in the goodwill of others not harm your interests when you are vulnerable to them" [52]. Gurung [24] proposed and validated a model for consumers' behavioral intention to engage in e-commerce transactions based on their privacy concerns, security concerns, trust beliefs, risk perception, attitude, and subjective norms and perceived behavioral control. Shin [46] developed a model for behavioral intentions to use social network services with the following antecedents: perceived security, perceived privacy, trust, and attitude. A literature review of empirical studies on individuals' privacy concerns and intention to use e-commerce was also recently presented in [33].

The models exhibit considerable differences, especially regarding the constructs included in them, but they also display remarkable similarities. These derive largely from a shared behavioral economics theory of *privacy calculus* assessment to the effect that when disclosing data in exchange for some economic or social benefit, data subjects tend to balance the costs and benefits of sharing or hiding personal information by controlling flows of their information [1], [8], [32]. Based on this privacy calculus, data subjects decide whether they are willing to disclose their personal information for certain benefits, or whether they will withdraw the information due to the expected negative consequences of the disclosure. It should be remembered that calculus is not always straightforward: Behavioral economics research such as that of Acquisti and Grossklags [1] has discussed the effect of information asymmetries and bounded rationality in engaging in privacy calculus. A data subject does not necessarily know when another entity has gained access to his/her personal information or used it, or may not know the potential personal consequences of such intrusions. The complexity of evaluating privacy costs and benefits complicates privacy decision making when data subjects, for example, have to consider multiple layers of outcomes and associated probabilities. Furthermore, privacy-related benefits and costs are rarely monetary and are often immaterial, thereby making privacy calculus more difficult to carry out. And finally, in our case of open traffic data, with unforeseen potential uses, the calculus gets subject to excessive uncertainty. Hence, for empirical purposes, we should also look beyond *privacy calculus* studies.

We reviewed and analyzed multiple models in order to summarize the existing knowledge on information privacy in the context of personal driving data collection for open traffic data. The searches were conducted using the following databases: Academic Search Premier (EBSCO), ACM Digital Library, Emerald, Lecture Notes in Computer Science, ScienceDirect (Elsevier), and Scopus. We searched for models specifically dealing with data subjects' willingness to disclose their data for various purposes, and the benefits gained from data disclosure. Our original inclusion criterion was that the models to be reviewed should include users' willingness to disclose data as the privacy behavior outcome, either directly or as an obvious part of service usage (such as the intention to use a personalized service, which in turn demands for personal data disclosure, as a dependent variable). In addition, models with users' payoff from data disclosure as the dependent variable were considered relevant to the study. Also behavioral economics studies dealing with e.g. incomplete information, bounded rationality and behavioral biases, and their influence on privacy decision making, were included in the review. Finally, in line with the reasoning above for going beyond *privacy calculus* models, we included one qualitative study on social and psychological issues underlying consumers' privacy concerns in the e-commerce exchange context. It was chosen because of its focus on how consumers increasingly perceive information as a commodity.

Below a short description of each of the selected 10 studies is presented. Following this, the key characteristics of these studies are summarized, including the context of the study, description of the data used, and suggestions for privacy management based on the results of each study (Table 1). Finally, at the end of this section, a synthesis of the main factors affecting data subjects' willingness to disclose their data is presented.

3.1 Description of Privacy Behavior Models

The study by Culnan and Armstrong [8] looked at consumers' willingness to disclose their personal information in creating profiles for marketing use and discovered whether their concerns about information privacy were addressed by fair information practices. The dependent variables of the research were as follows: 1. use without fair information practices (i.e., willingness to have personal information used for profiling without explicitly being told that fair information practices would be employed) and 2. use with fair information practices. The following independent variables were used: 1. behavior that indicated a concern for privacy and 2. prior experience with direct marketing. The results of the study showed that privacy distinguished people who were willing to be profiled from those who

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

were unwilling to be profiled only when people were not told that fair information practices would be followed. Therefore, the authors suggested that privacy concerns can be addressed by explicitly telling customers that the company observes fair information practices. In addition, according to the results, prior experience with direct or targeted marketing would discriminate among people who are willing to be profiled. When people were explicitly told that fair information practices are followed, only prior experience distinguished individuals who were willing to be profiled from those who were not.

In an extension of Culnan's and Armstrong's privacy calculus [8], Dinev and Hart [11] present a research model to study how the cumulative influence of Internet trust and personal Internet interest can outweigh privacy risk perception in the decision to disclose personal information in the context of online transactions. In the study, it was assumed that the salient beliefs influencing the intention to disclose the personal information in this context can be contrary, and that these beliefs together comprise a set of elements in a calculus (or decision process). The research model is composed of five constructs, as follows: 1. willingness to provide personal information to transact on the Internet (dependent variable), 2. perceived Internet privacy risk, 3. Internet privacy concerns, 4. Internet trust and 5. personal Internet interest (degree of cognitive attraction to Internet interactions). The results supported the researchers' hypotheses. The three factors most strongly related to the willingness to provide personal information were Internet privacy concerns, Internet trust, and personal Internet interest.

Chellappa and Sin [5] developed a model for predicting consumers' usage of online personalization as a result of the tradeoff between their valuing of personalization and their concern for privacy. In addition, consumers' intent to use personalization services was studied in relation to their trust in the vendor. The model consisted of two main constructs, as follows: 1. the value of online personalization (such as the fit that a product or service provides, the convenience of having it delivered in a proactive fashion, or personalization of product offerings) and 2. consumer concern for privacy. The results of the study showed that consumers' valuing of a personalized service is independent of their privacy concerns, and this positively affects their decision to use personalization services. Furthermore, the service usage is negatively affected by concern for privacy. In the model, consumers' valuing of personalization was almost two times more influential than the consumers' concern for privacy in determining usage. Therefore, in addition to not ignoring privacy concerns, vendors should focus on improving the quality of personalized services. Furthermore, the findings showed that understanding and evaluating the different values that consumers may place in enjoying various types of personalization is of critical importance. It is also possible for vendors to indirectly affect consumers' privacy concerns through trust building, and in this way, improve their ability to acquire and use consumer information. It is worth noting that consumers were concerned not just about their personally identifiable information, but also their anonymous and personally unidentifiable information.

The study by Liu et al. [34] focused on how a privacy-friendly business model should be designed in the context of location-based mobile services with a view to maximizing the payoff to the user. The empirical data consisted of a scenario-based survey whose respondents were categorized into two groups—risk-averse and risk-neutral users. The theoretical model used in the study was composed of the following four constructs: 1. personal data disclosed, 2. user's payoff, 3. personalization available and 4. control over user personal data. The results of the study showed that the extent of data disclosed has a significant negative effect on user payoff, whereas service personalization has a significant positive effect. Furthermore, control over user personal data has a positive effect on user payoff. It was also observed that the overall explanatory power of the model was increased within the group of risk-neutral mobile users. The authors suggested that service providers should take into account two different customer segments and design different value propositions for them: Risk-neutral users seek personalized service, while risk-averse ones seek data control. Based on the results of the study, they described a business model where, for example, the risk-neutral users get the services for free, and risk-averse users, who are more likely to pay to get the service, would subsidize the controls offered to the risk-neutral users.

Olivero and Lunt [43] presented a longitudinal, qualitative study on social and psychological issues underlying consumers' privacy concerns in the context of e-commerce exchanges. The results were based on grounded theory analysis of semi-structured interviews, follow up interviews, and integration of the analysis results with findings in the literature. According to their analysis, perceived risk and awareness of information collection/extraction are associated with a shift in concerns from issues of trust to issues of control. The results suggested that due to the awareness of the commercial uses of personal information, consumers increasingly perceive information as a valuable commodity. The interviewees showed willingness to disclose their personal information provided that their perceived benefits could justify the costs, such as time consumption and risks of vulnerability. In previous research, it has been shown that companies can build trust relationships with consumers by providing control over information. The results of this study suggest that by providing control to consumers, it is possible to address perceived risks and encourage consumers to exchange information that would not, based on trust, necessarily lead to a relationship. The authors stated that there is a need for instruments that allow consumers to make informed decisions in exchanges with companies and trade appropriate benefits.

The study by Son and Kim [48] looked at Internet users' information privacy-protective responses (IPPRs), with the aim of identifying determinants of these behavioral responses. The following antecedents of IPPRs were included in the nomological model developed in the study: 1. information privacy concerns, 2. perceived justice and 3. societal benefits from complaining. The study showed that, for online companies, it is essential to focus on perceived justice to convince Internet users to disclose correct personal information to them. Furthermore, they recommended that

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

online firms should increase their customers' perceptions of fairness by offering major benefits in return for releasing personal information and making sure that customers were aware of this (i.e., distributive justice), implementing fair procedures to protect the information and disclosing these procedures (i.e., procedural justice), and being trustworthy and honest in dealing with information privacy (i.e., interactional justice).

Malhotra et al. [35] presented a study on the influence of Internet users' information privacy concerns on consumers' decisions to release personally identifiable data in certain situations. The study conceptualized Internet users' information privacy concerns as the degree to which an Internet user is concerned about online marketers' collection of personal information, the user's control over the collected information, and the user's awareness of how his/her information is used. The findings of the study showed that online consumers consider it most important to be aware of, and have direct control over, personal information that is stored in marketers' databases. Based on these results, the authors proposed that the consumers should be able to easily check what type of information is collected, whether the information is correct, and how the information is used both in and outside the organization. Consumers should also have the opportunity to control the information in the organization's database by adding, deleting, and modifying it at will. In addition, the results of the study showed that trust in a marketer can significantly mitigate perceived risk and a customer's reluctance to release her personal information. Therefore, it is essential to aim at understanding how to increase customers' trust in their firms' handling of personal information.

The objective of Phelps et al.'s [44] study was to identify the types of personal information that consumers are most and least willing to provide to direct marketers and other retailers. Based on the results of the study, the authors emphasized the importance of addressing concerns related to the type of information collected and the amount of control that consumers have over subsequent dissemination of their information. The findings showed that consumers are least willing to provide financial and personal identifier information. It was also found that almost all consumers desire more information control, even those who are relatively unconcerned about data collection and use by marketers.

Jentzsch et al. [26] empirically studied consumers' decisions related to disclosure or non-disclosure of personal data in relation to purchase transactions. Based on their theoretical model with two service providers, the authors found that if the data practices of these service providers are difficult to compare, the decision of the consumer is not necessarily influenced by the terms of trade for personal data and the consumer tends to ignore them due to their complexity. In this situation, the service provider is not able to obtain a competitive advantage through privacy settings that could be used to fit consumer preferences. Correspondingly, the authors observed that if service providers' offers are placed close to each other and it is possible for the consumers to compare the amount of data that are collected within the transaction, they take information practice into account in their decision making.

In their study, Spiekermann et al. [49] conducted an experiment on university students to discover data subjects' actual disclosing behavior during an online shopping episode. They found that, regardless of reporting specific privacy concerns, most of the participants did not live up to their privacy preferences. Specifically, participants clustered into the groups *Profiling averse* and *Privacy fundamentalists* (the latter being most concerned about all aspects of privacy) did not behave according to the attitude they had expressed. It was observed that the participants were willing to reveal private and even highly personal information when communicating with the anthropomorphic 3-D bot. Based on these findings, the authors recommended that privacy technologies be designed in a way that enables even moderately computer-literate data subjects to protect themselves from the degree of self-disclosure they are afraid of.

Table 1 summarizes the key characteristics of the reviewed studies, including the context of the study, description of the data used, and suggestions for privacy management based on the results.

3.2 Synthesis of the Main Factors Affecting Data Subjects' Willingness to Disclose Data

As a conclusion from the literature review, we would like to highlight the main factors affecting data subjects' willingness to disclose data. Data subjects' information privacy concerns negatively affect their willingness to disclose data directly, through a mediator effect, or as a mediator [5], [8], [11], [35], [48]. For example, in the model in [35], Internet users' information privacy concerns were mediated by trusting beliefs and risk beliefs. According to these results, trust can significantly mitigate perceived risk and customers' reluctance to release their personal information. In contrast, according the model in [11], Internet privacy concerns are positively dependent on perceived internet privacy risk. Here, Internet trust is considered as a mediator of perceived Internet privacy risk, independent of privacy concerns. Further, in [48], the negative effect of perceived justice related to the refusal to provide information is presented as a separate construct, independent of information privacy concerns. Users' cost-benefit considerations bring out the following factors with a positive effect on willingness to disclose data: offering customers benefits in return for releasing personal information (in addition, it has to be ensured that customers are aware of these benefits) [48], justifying costs as time consumption [43], and demonstrating the value of personalization [5], [34].

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 71-89

This paper is available online at www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

Table 1: Key characteristics of the reviewed studies

Study	Context of the study	Description of data	Suggestions for privacy management
Culnan and Armstrong (1999)	Use of personal information gathered from prospective subscribers to interactive home information services; using personal information for targeted advertising based on customer profiles that are compiled by the interactive service providers	Random sample of 1,000 respondents, collected in 1994 in the US	Informing data subjects explicitly about the observance of fair information practices
Dinev and Hart (2006)	E-commerce transactions	Sample of 369 respondents from the US	Seeking ways of building data subjects' confidence and minimizing their privacy risks proactively
Chellappa and Sin (2005)	Online personalization within the following online industry categories: personal computers, automobile, apparel, financial services, travel services	Survey with 243 responses from consumers who had purchased online products or services	Improving the quality of personalized services; Understanding and evaluating data subjects' values concerning various types of personalization; Affecting data subjects' privacy concerns indirectly through trust building; Worth noting: data subjects may also have concerns about their anonymous and personally unidentifiable information
Liu et al. (2011)	Location-based mobile services	187 undergraduate students	A business model is described where, for example, the risk-neutral users get the services for free, and risk-averse users subsidize the controls offered to the risk-neutral users
Olivero and Lunt (2004)	E-commerce exchanges	23 semi-structured interviews (repeated e-mail exchanges), five face-to-face follow-up interviews	Providing control to data subjects in order to address perceived risks and to encourage them to exchange information; Establishing instruments that allow data subjects to make informed decisions in exchanges with companies and to trade appropriate benefits
Son and Kim (2008)	Internet users' IPPRs in the general context of online companies	Survey with 541 responses from panel members of a market research firm	Offering data subjects major benefits in return for releasing personal information and making sure that the data subjects are aware of the benefits; Implementing fair procedures to protect information, and disclosing these procedures; Being trustworthy and honest in dealing with information privacy
Malhotra et al. (2004)	Consumers' decisions to release or not to release personally identifiable data in e-commerce	742 household respondents from two separate surveys	Making sure that the data subjects can easily check what type of information is collected, whether the information is correct, and how the information is used both in and outside the organization; Providing data subjects with the possibility of controlling the information in the organization's database by adding, deleting, and modifying it at will; Understanding how to increase data subjects' trust in handling personal information
Phelps et al. (2000)	Direct marketing; exchanging personal information for shopping benefits	556 responses from two sampling frames (known and recent catalog shoppers, randomly selected residences)	Providing data subjects with control over subsequent dissemination of their information; almost all data subjects desire more information control, even those who are relatively unconcerned about data collection
Jentzsch et al. (2012)	Consumers' decisions on disclosure or non-disclosure of personal data in an Internet purchase transaction	443 responses from students who participated in a laboratory experiment	Establishing privacy settings based on consumer preferences and ensuring that the data subjects are able to compare the data practices of a service provider to the ones applied by other service providers
Spiekermann et al. (2001)	E-commerce transactions in an environment where an anthropomorphic 3-D shopping bot involved users in a sales dialogue and gave product recommendations	200 observations from a laboratory experiment (95% of the participants were students)	Designing privacy technologies that enable data subjects to protect themselves to the extent they consider appropriate

In addition to privacy concerns as such, data subjects' willingness to disclose their data depends on several other factors. Willingness to disclose data can be managed by controlling some of these factors. These factors and their effects on willingness to disclose data are categorized to six aspects (informing, trust, benefits, information type, control, individual factors), as summarized in Table 2.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

Table 2: Factors affecting data subjects' willingness to disclose data

Factor	**Effect on willingness to disclose data**
Benefits: Data subjects' valuing of a personalized service affects their decision to use personalization services [5] Perceived benefits that justify the costs, such as time consumption and risks of vulnerability [43] Offering data subjects major benefits in return for releasing personal information [48], for example, by providing personalized services [5]	 positive positive positive
Trust: Trust in a delivery channel such as the Internet [11] Trust in a marketer [35] Prior experience with data collection or utilization [8]	 positive positive positive
Informing: Data subjects' awareness of privacy practices [35] Data subjects' awareness of the use of fair information practices (due to informing data subjects) [8] Implementing fair procedures to protect the information, and disclosing these procedures [48]	 positive positive positive
Control: Control over personal information provided [35] [43] [44]	 positive
Information type: Sensitivity of requested information [35]	 negative
Individual factors: User type: risk-averse/risk-neutral [34] Personal Internet interest [11]	 negative/positive positive

The review revealed the significance of fair information practices and informing the data subjects about them [8]. Service providers should implement fair procedures to protect the information, disclose these procedures (i.e., procedural justice), and be honest in dealing with information privacy (i.e., interactional justice) [48]. Consumers should be able to easily check what type of information is collected, whether the information is correct, and how the information is used both in and outside the organization [35]. Consumers should also be able to control the information in the organization's database by adding, deleting, and modifying it at will [35]. Correspondingly, within the driving data collection, how important the data subjects consider the ability to control their data disclosure should be investigated, for example, by switching the on-board device off or by removing the data from the driving data operator's system.

Trust has a significant effect on willingness to provide personal information [11]. It seems that people with prior experience in activities related to data collection and usage are more willing to disclose their data compared to people with no prior experience [8].The benefits of revealing private information are also of great importance. According to Chellappa and Sin [5], consumers' valuing of a personalized service is independent of their privacy concerns, and positively affects their decision to use personalization services. It seems that people are willing to disclose personal information provided that the perceived benefits justify the costs, such as time consumption and the risk of vulnerability, as they learn to understand the disclosed information as a commodity of value [43]. It is recommended that customers be offered major benefits in return for releasing personal information; moreover, it should be ensured that customers are aware of them [48].

4 Empirical Study on Driving Data Related Privacy Concerns

It was shown in the previous section that several studies have proposed privacy behavior models and presented empirical results on the effects of privacy concerns on persons' intention to use different Internet services and e-commerce solutions. Such research has demonstrated information practices, trust, benefits, and control over data are important factors in users' willingness to disclose their private data. However, the previous literature has neither addressed personal driving data privacy behavior nor the production of open data as a result of personal data collection. Usually, earlier studies have assumed—at least implicitly—that the service users are not exactly aware of what data is collected and for what purposes, and that data collectors' data usage is obscure or not objectively defined and known.

As our interest lies in driving data and their potential, unforeseen usage in future traffic service innovations, we have conducted empirical studies in a real driving data–based service context. Here, we present the empirical results of two driving data–based service pilot studies, where we focused on data subjects' privacy concerns and their willingness to disclose personal driving data.

The first driving data–based service pilot reported in this paper is a *driving data multiservice system* (DMS) pilot. This DMS system collected data similarly to what has been suggested by the Ministry of Transport and Communications

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

of Finland in its plans for kilometer-based taxation. In return for a monthly fee, the suppliers offered the customers such services as an automatic driver's log, fleet management, vehicle routes, and driving style information. This service required the driving data operator company to collect, manage, and process customers' driving data. The users of the service included private users, private entrepreneurs or proprietors, and company users. The second smaller pilot, which complements the results of the DMS pilot study, was based on driving data collected by a fleet of taxis in a city. Here, professional driving data was collected through on-board devices that were provided by a company acting as a data operator. The collected data, which also included other sensor data from the car in addition to the location of the vehicle, was intended for use in the production of new innovative intelligent traffic services.

4.1 Research Setting

The empirical study was conducted in four stages. At first, the DMS research was conducted in three stages utilizing a multi-method approach that comprised stakeholder interviews, user interviews, and a user survey. After this, the results were complemented by an interview of the taxi drivers whose driving data were collected through on-board devices. These four stages and the related empirical data sets are presented in Table 3.

Table 3: Empirical datasets

Data collection	**Stage 1: DMS stakeholder interviews**	**Stage 2: DMS user interviews**	**Stage 3: DMS user survey**	**Stage 4: Taxi driver interviews**
Subjects	Representatives of service stakeholder organizations (service providers, data users, and public authorities)	Service users and non-users (non-adopters and dropouts)	Service users	Taxi drivers participating in driving data collection
Research method	Open theme interviews	Semi-structured theme interviews	Survey	Semi-structured theme interviews
N	10	11 users, 7 non-users	62	3

In the first stage, 10 representatives of DMS stakeholder organizations, including public authorities and companies with experience with large systems involving personal data collection and utilization, were interviewed on their views about personal data privacy. In addition, parties that were not currently participating in the pilot system were included in as interviewees. The interviews were conducted as open theme interviews, either in person or via telephone. The results from this stage were utilized in planning the DMS user interviews in the second stage.

In the second stage, 11 service users were interviewed utilizing semi-structured theme interviews via phone to discover their views on personal data privacy. These interviews were planned to take about 30 minutes, and the actual duration of the interviews varied between 13 and 47 minutes (31 minutes on average). All interviews were recorded and transcribed by the interviewers. The interview summaries were cross-checked by another interviewer to ensure consistency with the original data. In addition to the actual service users, seven non-users were interviewed. Some of these persons had considered adopting the service but decided not to take it into use, and some had quit using the service. The non-users were also interviewed via phone, except for one person, who volunteered to be interviewed face-to-face. Semi-structured interviews were planned to be 20 minutes long, and the actual duration of the interviews varied between 12 and 41 minutes (19 minutes on average). All interviews were recorded, and after the interview, data were arranged in a tabular form.

In the third stage, a user survey was designed based on the existing information privacy concern measurement instruments presented in the literature [35], [47], [50], legislation on data privacy, the results of the user interviews, and the identified possible benefits of the DMS services for the users. The survey was conducted as part of a more extensive service adoption and data privacy survey. The survey included questions on users' willingness to disclose their personal driving data to gain various benefits, namely compensation for service costs, better service, personalized service, services provided by different parties, and advanced traffic information services for common use. In terms of the users' data privacy concerns, the respondents were asked about whether they had had concerns in the service adoption stage about the purposes of using their data and about the parties to whom the data would be disclosed. Respondents were also asked about their personal concerns about information protection issues to assess the difference between users' actual privacy concerns and their theoretical consideration of various privacy issues discussed in public. The survey was made available online; a request to participate in the survey was sent to the service users by e-mail, including a link to the questionnaire. The size of the DMS user group was roughly 500 people. In three weeks, 62 responses were obtained; 16 of the respondents were private users and 46 were company users. Half of the respondents used the service daily, one third used it weekly, and the rest monthly or more seldom. Of the respondents about 81 % were males and 19 % females.

In the fourth stage, carried out in the taxi data collection pilot, interviews were conducted with taxi drivers. Three taxi drivers were interviewed as a group to assess taxi drivers' attitudes towards collection of their driving data, and for identifying their possible data privacy concerns stemming from onboard device data collection. These devices were installed in eight vehicles approximately six months before the interview.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

4.2 Results of the DMS Stakeholder Interviews

The DMS stakeholder interviews were carried out in the beginning of the study, in order to understand the distinct aspects of privacy related to driving data, from the points of view of different DMS stakeholders. In these interviews, various data subjects' privacy concerns were identified. Data subjects do not always have enough information about what personal data is collected, what purposes it is utilized for, and to what parties it is disclosed. With respect to data disclosure to third parties, the data subjects have questions concerning data protection practices and management systems, as well as form and content of the data disclosed to third parties. Considering the systems of the data controllers and processors, the stakeholders reported that data subjects appear to have concerns about who have authorized access to the data. The results from the stakeholder interviews were utilized in planning the DMS user interviews in the next stage of the study.

4.3 Results of the DMS User Interviews

In the DMS user interviews, users' and non-users' information privacy concerns were investigated as regards data collection and processing, data disclosure to third parties, and access to the driving data by various parties. Further, the relation of the non-users' privacy concerns to their service adoption decision was studied, as well as users' awareness of the file description and the service terms and conditions, and their information needs in respect of data collection and privacy protection.

In the service adoption stage, most of the interviewed users (7/11) did not have privacy concerns or only had minor concerns (e.g., one user had pondered whether somebody has access to his/her information but he/she still thought that the information was not very useful for anybody and was not worried about the issue). The users' privacy concerns were related to, for example, data disclosure to third parties, authorized data access, and unauthorized data access. In other words, they had been thinking about who had access to their data, where the data were released, the purposes for which the data were used, and whether the authorities monitored users' speed to impose fines.

Four out of seven non-users reported they had pondered information privacy issues when making the adoption decision. They had read critical comments in magazines about the privacy problems related to services of this kind, and had discussed the topic with their friends. Despite this, none of non-users considered information privacy concerns as critical to their adoption decision or problematic in the use of the DMS. Instead, the main reason for not adopting or quitting was a too-high price compared to the value added (such as less effort in keeping a driving diary and reporting, information that helps to improve driving habits and decrease fuel consumption, or willingness to contribute to studies of this kind where new services are being developed). Some non-users also reported that they applied alternative means of keeping a diary.

We observed that not all of the users were aware of those to whom their driving data was disclosed or for what purposes it was used. It was also found that, quite often, there seems to be lack of factual information on data privacy and personal data protection. Although this information was available, it may be presented in a way that is difficult to read for the users. The users presented several specific information needs concerning the processing of personal data, as follows: what purposes vehicle position and user driving behavior information is used for, who is following driving behavior information, how the data are stored, to whom the data are disclosed, whether the data are disclosed to other companies or public authorities, who can access the data and in what form the data is accessed (e.g., detailed data or averages). One interviewee suggested that users should be informed if there are plans to use this information for marketing purposes or when an external party has access to the information. In this way, users would be able to decide whether to continue using the service. External validations of the service were also proposed to enable informed decisions by users when adopting the service.

Some users reported that they had been thinking about value of their data and benefits gained from disclosing data to be used by third parties. One of these users stated that benefits (such as financial benefits) from data disclosure are decisive in terms of the user's willingness to disclose his/her information, and one mentioned that he/she found it strange to think that somebody could gain financial benefit from users' data in a system whose use is paid for by the user him/herself. The interviewees also stated that the users should themselves have the opportunity to control usage of their data.

The interviewees did not report serious distrust. Many of the users also perceived that driving data are not sensitive. If the interviewee did not report any privacy concerns and revealed that she/he trusted some company involved in the service provision, she/he was asked how this trust had arisen. Since the DMS service was marketed by a bank and insurance company, which users perceived as a trusted organization, they trusted it. Some users reported that they trusted the company installing the telematic on-board devices in the car, having had a long customer relationship with this company. This implies that trust can be transitive, that is, when there was one highly trustworthy company involved in the service provision, the interviewees were able to trust the whole network. Two out of 11 users and all non-users (7 interviewees) reported they felt this kind of a transitive trust.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

In summary, the interview findings revealed that, from the users' point of view, the following key factors significantly affect the service users' privacy concerns related to their driving data: communication, transparency of processing and data disclosure; trust of companies in the business network providing the personal data–based services; and users' control over their personal driving data. These empirical results in the context of driving data correspond well with the synthesis of the earlier privacy behavior model analysis presented in the previous chapter. The factors brought up in the interviews were essentially the same. The value of their own data had also been considered by the drivers, leading to thinking about receiving benefits as a compensation for disclosing personal data for purposes other than their own service provision.

4.4 Results of the DMS User Survey

In the DMS user survey, users' willingness to disclose their data for gaining various benefits and their privacy concerns related to data disclosure were investigated. Users' willingness to disclose their data for benefits was discovered in the survey through a multiple choice question with the following benefits: compensation for service costs, better service, personalized service, services provided by different parties, and advanced traffic information services for common use. The results of this survey item are presented in Table 4. In addition to the willingness to disclose data for specified benefits, the acceptance of data disclosure was calculated from the responses, indicating willingness to disclose driving data to gain at least one of the listed benefits.

Table 4: Respondents' willingness to disclose their driving data to gain benefits

Benefit	Respondents (%)
Acceptance of data disclosure for at least one benefit	69.4
Compensation for service costs	46.8
Advanced traffic information services for common use	35.5
Better service	27.4
Services provided by different parties	22.6
Personalized service	19.4

The results of the survey showed that nearly 70% of the respondents would accept data disclosure to gain some benefits, whereas about 30% of the respondents would not be willing to disclose their data for benefits at all. Private users seemed to be remarkably more willing to disclose their data (93.8%) compared to company users (60.9%). When disclosing their personal data to gain benefits, the users seemed to be most willing to disclose their data for compensation for service costs and to obtain advanced traffic information services for common use. A substantial difference between the user groups was observed for the benefits *Compensation for service costs* (private users 81.3%, company users 34.8%) and *Advanced traffic information services for common use* (private users 75.0%, company users 21.7%).

Users' data privacy concerns were assessed using two 5-point scales (from *Not at all concerned* to *Very much concerned*). In the questionnaire, the users were asked to evaluate how concerned they had been, in the service adoption stage, about what purposes their data would be used for, and to whom the data would be disclosed. The levels of these data privacy concerns are presented in Table 5. Moreover, Table 6 presents some descriptive statistics calculated from the responses.

Table 5: Levels of users' data privacy concerns

Level of concern	What purposes the data are used for (%)	To whom the data are disclosed (%)
Not at all (1)	12.9	8.1
Very little (2)	11.3	11.3
Little (3)	35.5	27.4
Much (4)	19.4	22.6
Very much (5)	19.4	27.4
No response	1.5	3.2
Total	100.0	100.0

Of the respondents, 19.4% reported that they had much concern and 19.4% reported that they had very much concerns about what purposes the data would be used for (one respondent did not give a response to this item). Moreover, 22.6% reported that they had much concerns and 27.4% reported that they had very much concern about those to whom their data would be disclosed (two respondents did not give responses to this item). For the statement *What purposes the data are used for,* the distribution of the responses had no strong tendencies, except for the category *little,* which represented 35.5% of the responses.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

Table 6: Statistics for responses to the privacy concern statements

Survey statement	Mean	Median	Mode	Std. deviation
What purposes the data are used for	3.21	3	3	1.266
To whom the data are disclosed	3.52	4	3 and 5	1.255

Based on the statistics, it can be observed that for the responses to the statement *To whom the data are disclosed*, there is a difference of 0.5 between the mean and median. In addition, there is a double mode of 3 and 5. This indicates that the distribution is negatively skewed and the respondents tended to be more concerned with privacy.

We aimed to distinguish users' actual data privacy concerns from just pondering about privacy matters discussed in public. Therefore, respondents were also asked about their personal concerns about information protection issues through the statement *Did you personally find data protection issues problematic?* Levels of users' actual privacy concerns are presented in Table 7.

Table 7: Levels of users' actual data privacy concerns

Level of actual data privacy concerns	Respondents (%)
Not at all (1)	24.2
Very little (2)	24.2
Little (3)	27.4
Much (4)	6.5
Very much (5)	14.5
No response	3.2
Total	100.0

In contrast to the results presented above, it seems that users may be pondering various information privacy issues but are still not necessarily personally worried about their information privacy. For example, users may have been following the privacy protection discussion in public, but are not greatly concerned about how their actual data are managed. Of the respondents, 24.2% reported that they considered data privacy issues not at all problematic and 24.2% reported very little concern about data privacy issues being problematic (two respondents did not complete this item). Consequently, nearly 50% of the respondents reported no personal privacy concerns or that they had very little concern.

On the whole, the empirical results from this study showed that users' privacy concerns do not necessarily hinder the adoption of personal data–intensive services or personal data disclosure. If the provided services are valuable to them, or some benefits are offered, it seems that most users have some willingness to disclose their data.

4.5 Interview with the Drivers of the Taxi Pilot

In the interview with the taxi drivers participating in the driving data collection pilot, drivers' attitudes towards data collection were investigated. This interview aimed to identify drivers' possible data privacy concerns stemming from onboard device data collection. It also aimed to discover drivers' awareness of data collection and data privacy protection as well as their related information needs.

The interviewees did not perceive any problems with driving data collection, provided that no extra costs or effort would be required on their part. The interviewees stated that they had positive attitudes toward data collection and its utilization for research purposes. However, they did not find the devices or the data collected by them particularly useful or interesting, as the same type of information is available directly from their vehicles.

The interviewees did not report any data privacy concerns and they stated that they had not been pondering privacy matters during the use of the device. According to the interviewees' perception, the data were processed appropriately and not disclosed unless permitted by the data subjects. In general, the interviewees did not see the collection of driving data as problematic because location data are already collected extensively (e.g., by mobile phones). Moreover, the interviewees had not observed any reluctance to participate in the data collection among their colleagues. One of the interviewees suggested that taxi drivers are quite willing to participate in activities that aim to improve traffic conditions. The interviewees also mentioned their personal long-term experience with data collection within the taxi system as an antecedent to their trust in appropriate data processing. However, they noted that if the data collection were carried out by any parties unfamiliar to the drivers, their trust could decrease. The interviewees stated that an adequate data privacy level can be reached by anonymizing the data.

Despite the low data privacy concerns, the interviewees thought that it would be reasonable to discuss data protection matters with the taxi drivers when they started to use the data collection devices. Through such a measure,

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

drivers would be provided with information on how privacy protection had been taken into account and managed in the data collection. According to the interviewees, it is essential to inform the drivers of the identity of the data controller and the data types that the controller can access. In addition, the accuracy of the data and its anonymization should be highlighted. The interviewees also stated that trust can be diffused in an organization if a person who is already familiar with the data collection activities encourages the rest of the personnel, for example, to start using data collection devices. In addition, the information should be presented orally in a popularized form, easy to understand, so that it is explicitly stated that a single driver is not distinguishable and that irrelevant parties do not have access to the data.

4.6 Summary of the Findings

As a summary of our empirical studies on personal driving data of the multiservice system DMS, we can conclude the following. *Most data subjects had privacy concerns*. The privacy concerns are most often related to *disclosing driving data to third parties, then to stakeholders with authorized data access, and least often to unauthorized data access*. To illustrate the point, data subjects are especially doubtful about potential disclosure of speed data to the police, or location-based data for marketing purposes. However, even though data subjects may ponder various privacy issues (e.g. after following the privacy protection discussion in public) *they may not worry about their personal information privacy at all*.

Against this backdrop, an important finding is, especially in terms of designing of new services, that the large majority of data subjects (69% in the DMS survey) are *willing to disclose their driving data for benefits*. Especially among private users, willingness to disclose data for gaining benefits seems to be high. It also seems that some users are thinking about value of their data, and *in return to data disclosure, they expect to receive some benefits*. This observation is supported both by the previous research [43] and the results of the DMS user interviews. The benefits can be self-serving, such as economic benefits or better personal services, or altruistic, for contributing to better services for all. Furthermore, the DMS user interview results showed that, instead of privacy concerns, low added value of the service perceived by a user compared to its price was the main reason for not adopting or quitting the service. This may imply that the net value of the service takes precedence over users' privacy concerns.

The results of the DMS user interviews also revealed a trust factor. The data subjects exhibited different levels of trust for different organizations. For example, a bank and insurance company as a part of the ecosystem is considered highly trusted organization. Some users had had a long customer relationship with the company installing the telematic on-board device, and hence they trusted this party in the ecosystem. Furthermore, the transitiveness of trust was also observed. If the data subject trusted, for example, one organization in the ecosystem, or even one person in the service delivery chain, he/she would trust the whole system. Thus, the reputation and experienced trustworthiness of the driving data operator's and the whole associated ecosystem's organization is important. Similar to the existence of data privacy concerns, personal long-term experiences of data collection within certain contexts may be an antecedent to trust in appropriate data processing.

Our study confirms also the earlier findings in the literature on the importance of informing the data subjects about privacy issues. Especially the interviewed users wanted more information on the identity of the controller of the private data, purposes of the data usage, how the data are stored, whether the data are disclosed to third parties, in what form the data are accessed, accuracy of the accessed data, and how privacy protection has been preserved in data collection. According to the present – and most likely to the future – legislation on private data, this information is needed to properly inform users for adopting new open data services and for informed consent to use their information. Our conclusion is that in the DMS context Privacy by Design, or a similar approach, is a crucial antecedent to opening the data for alternative uses. Below we shall elaborate this claim further.

5 Application of the Results

The empirical results on data subjects' privacy behavior reported in this paper are building on the prior research, but primarily descriptive in nature. This is because we had access to our data in multiple stages during the development of a driving data multiservice system (DMS). This way, we also had an opportunity to look beyond the most common *privacy calculus* models, and we are in a position of reflecting upon the design and launch of similar new personal data collection initiatives for producing open data for service innovation.

To be able to collect personal driving data from a target group of drivers in a legitimate way, the data operator needs to have an agreement on data collection with the data subjects. Following the data protection laws and guidelines, this is accomplished through informed consent. Preparing *adequate informed consent* that guarantees maximum agreement from the data subjects means that the data operator must study the conditions affecting the data collection in the target group. The data subjects' point of view is vital here, as data collection depends on their subjective voluntary agreement. With this user-centric view, the data operator can understand the benefits of the planned service for the data subjects and the perceived personal data disclosure costs concerning felt privacy risk that affect the decision to give the consent before the actual consent is requested from individual drivers. In particular, data subjects' privacy concerns that may negatively affect the decision to give the consent, can be addressed in preparing the informed consent, thereby increasing the likelihood of a positive decision to allow data collection. In the

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

following, our results and their future application opportunities are briefly discussed from this user-centric perspective on driving data collection.

As data subjects often expect to receive benefits as a compensation for disclosing their personal data, how they value various benefits should be identified. Based on our empirical studies, we have found that, in addition to self-serving, *egoistic benefits*, many of the drivers see value in *altruistic benefits* of the service. They are willing to disclose their personal driving data for the common good. This is a significant finding in relation to the disclosure of data for open traffic data, which are meant to be used for the common good. This is an aspect that is also evident in many present-day car navigators, where tracking of location is requested based on users who are *willing to improve traffic fluency*. To properly identify the benefits of the planned service for the data subjects, the data operator should pay attention to both of these types of benefits.

Privacy risks perceived by a user may exist even if data subjects' privacy is not compromised and despite valid data protection regulation, fair information practices, and PbD guidelines for designing the data collection and management system being followed by the driving data operator. According to the results of the empirical part of this study, privacy concerns about data disclosure to third parties are the most relevant for discussion in the context of driving data–based open traffic data services (i.e. originally unforeseen, potential uses of private data). When evaluating data subjects' privacy concerns, it also has to be borne in mind that they may be pondering information privacy issues, for example, after following public discussions of privacy protection, but are not necessarily personally worried about their information privacy. As demonstrated in the empirical part of this study, to be able to deal with the difference between data subjects' actual privacy concerns and theoretical considerations of various privacy issues, their personal concerns about information protection issues have to be investigated. In order to evaluate data subjects' privacy risk perceptions, their privacy concerns can be measured utilizing the existing information privacy concern measurement instruments [35], [47], [50] and the results of the empirical research that was conducted in this study.

In addition to the context-related perceived privacy risks, there are *individual factors* affecting data subjects' willingness to disclose data. For example, our empirical results show that a user's *prior experience* with driving data collection seems to have an effect on willingness to disclose data, and such past experience may also mitigate perceived privacy risks. Similar to prior experience, as was observed when interviewing taxi drivers, it can be concluded that *personal altruistic interest* in developing traffic conditions could positively affect willingness to disclose driving data. One individual factor not yet extensively included in the previous research models is user type or user *personality traits*, beyond risk aversion. For example, Korzaan and Boswell [30] discussed the Big Five personality traits in their study on how personality traits and information privacy concerns influence behavioral intentions. Similarly, Junglas et al. [28] investigated the effect of the personality traits on privacy concerns in the context of location-based services. There are validated scales for measuring personality traits, such as the 10-item seven-point scale in [23], which was also used in [28].

Our results revealed the *crucial role of communicating well with the data subjects about data collection and processing* when they are considering whether to disclose personal data. Since data subjects may not have enough factual information on data privacy and personal data protection, their awareness of the driving data recipients and purposes of the data collection has to be identified. In addition, data subjects' potential information needs should be specified. In particular, it is important to provide the following information: identity of the controller, purposes of the data usage, how the data are stored, whether the data are disclosed to third parties, in what form the data are accessed, accuracy of the accessed data, and how privacy protection has been managed in data collection. Finally, it should be considered whether data subjects need to be provided with additional information on data collection due to their low level of experience with driving data collection or low interest in participating in data collection, or whether certain personality types should specifically be taken into account when informing the data subjects.

It is also of high importance to investigate *potential trust issues* related to driving data collection because of the role of such issues in the formation, occurrence, and mitigation of privacy concerns, as well as their effect on willingness to disclose data. For example, Jarvenpaa and Tractinsky [25] used a seven-item, 7-point scale to measure consumers' trust in an Internet store. Scales of this kind could be used to evaluate data subjects' trust in driving data operators, as well as separate organizations belonging to the ecosystem.

6 Conclusions

The objective of this paper was to study privacy issues in the context of service building for users' personal driving data in the production of open traffic data. The main findings of our empirical study on personal driving data of the multiservice system DMS were as follows:

1. Many users are concerned about privacy issues.

2. They are worried about potential data disclosure to third parties and the use of the data for purposes beyond the original ones, and both authorized and unauthorized data access.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

3. Despite these characteristics, users are willing to disclose their driving data for benefits.

4. The reasons for voluntary data disclosure can be of a selfish nature, like economic compensation and improved personalized service, or altruistic, like contributing to the development of services for all.

5. Users' awareness of the driving data collection and processing may not be sufficient from their point of view. Users' information needs should be identified and fulfilled.

6. The good reputation and trustworthy image of the service provider will mitigate privacy concerns.

7. Trust can be transitive: If a data subject trusts some organizations in the service provider's ecosystem, or even a specific person in the service delivery chain, he/she seems able to trust the whole ecosystem. This is an interesting finding for creation of new, innovative services requiring collaboration between several organizations, and would be worth investigating in more depth.

We recommend that service providers pay attention to maintaining trust by carefully informing the users about parties to whom their private data will be disclosed and for what purposes they will be utilized. This can reassure the users of the safety of adopting and continuing to use the services, and even positively affect their willingness to disclose their data. Together with possible external validation of the service, this information will enable informed decisions by users when they adopt the services.

Building on the findings above, it is necessary and possible to construct privacy-friendly data policies even for services requiring highly sensitive private data, such as tracking of a person's whereabouts and driving speed. It should be noted that personal data collection for other than the legitimate purposes of the public authorities or related to business is only allowed for specified reasons with data subject's informed consent. Personal driving data collection to produce and release open traffic data is clearly an activity that requires consent from the data subjects. It also implies the importance of the PbD [2] (or a similar method) in opening data for purposes beyond the original ones, because it covers not only information technology issues, but also business accountability and complex infrastructure relations in design. By taking advantage of the PbD guidelines, personal driving data can be collected, managed and processed in a *transparent manner* without compromising data subjects' privacy. There are many practices and methods available, such as privacy assessment frameworks [4] and PETs [27]

If there exists a driving data operator that follows valid data protection regulations, respects fair information practices, and designs the data collection and management system according to the PbD guidelines, personal driving data can be collected and open traffic data can be produced. We think that *the critical barrier of being able to collect any personal driving data is that an informed consent is needed from the data subject*. We suggest that by examining the data privacy aspects studied in this research, service providers can assess the view of their target group on critical issues related to data disclosure such as expected benefits (personal or altruistic), trust in the service provider network, and data subjects' information needs.

Considering the validity and generalizability of our present empirical findings, a couple of points should be noted. Despite our attempts to build our empirical studies on the previous research, and after having followed the evolution of a DMS empirically, the findings are based on a relatively small number of interviews (18) and survey responses (62). We therefore welcome further studies challenging these findings. There are still many ways in which we can further our understanding of designing service innovations on the basis of personal data–based open data. Both conceptual and comparative research is needed to provide more insight into alternative privacy-friendly business models. Of equal importance in research is to gather more empirical data on the challenges of and potential resolutions for privacy issues in real-life services.

Acknowledgments

The research presented here was carried out through the Data to Intelligence (D2I) research program funded by Tekes and a consortium of companies. Some of the empirical data used in this study was collected via the Cooperative Traffic research program, funded by Tekes and a consortium of companies.

Website List

Site 1: Digiroad
http://www.digiroad.fi/en_GB/

Site 2: Digitraffic
http://www.infotripla.fi/digitraffic/

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

Site 3: EZPass
http://www.e-zpassiag.com/

Site 4: FasTrak
https://www.bayareafastrak.org

Site 5: Google Maps
https://maps.google.com/

Site 6: Helpten
http://www.helpten.fi/en/

Site 7: New York City Traffic Map
http://nyctmc.org/

Site 8: OnStar
https://www.onstar.com

Site 9: OpenStreetMap
http://www.openstreetmap.org

Site 10: Oulu City Traffic Map
http://www.oulunliikenne.fi/#/joukkoliikenne

References

[1] A. Acquisti and J. Grossklags, What can behavioral economics teach us about privacy? in Digital Privacy: Theory, Technologies and Practices (A. Acquisti, S. De Capitani di Vimercati, S. Gritzalis, and C. Lambrinoudakis, Eds.). Boca Raton, FL: Auerbach Publications (Taylor and Francis Group), 2007, pp. 363-377.

[2] A. Cavoukian. (2009, January) Privacy by design ... Take the challenge. Information and Privacy Commissioner of Ontario. [Online]. Available: http://www.privacybydesign.ca/content/uploads/2010/03/PrivacybyDesignBook.pdf

[3] A. Cavoukian. (2012, December) Operationalizing privacy by design: A guide to implementing strong privacy practices. Information and Privacy Commissioner of Ontario. [Online]. Available: http://www.ipc.on.ca/images/Resources/operationalizing-pbd-guide.pdf

[4] Center for Internet and Society. (2011, November) PET cyber wiki. Center for Internet and Society. [Online]. Available: http://cyberlaw.stanford.edu/wiki/index.php/PET

[5] R. K. Chellappa and R. G. Sin, Personalization versus privacy: an empirical examination of the online consumer's dilemma, Information Technology and Management, vol. 6, no. 2-3, pp. 181-202, 2005.

[6] H. Chesbrough, Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era. San Fransisco, CA: Jossey-Bass, 2011.

[7] H. Chesbrough, W. Vanhaverbeke and J. West, Open Innovation: Researching a New Paradigm. Oxford: Oxford University Press, 2006.

[8] M. J. Culnan and P. K. Armstrong, Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation, Organization Science, vol. 10, no. 1, pp. 104-115, 1999.

[9] E. De Cristofaro and M. Wright, Eds., Privacy enhancing technologies, Proceedings of 13th International Symposium, PETS 2013, Bloomington, 2013.

[10] E. den Ouden, Innovation Design: Creating Value for People, Organizations and Society. London: Springer, 2012.

[11] T. Dinev and P. Hart, An extended privacy calculus model for e-commerce transactions, Information Systems Research, vol. 17, no. 1, pp. 61-80, 2006.

[12] EC COM. (2012, January) Proposal for a regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General data protection regulation). European Commission. [Online]. Available: http://www.europarl.europa.eu/document/activities/cont/201305/20130508ATT65776/20130508ATT65776EN.pdf

[13] EC. (1995, November) Directive 95/46/EC (Data protection directive), Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. EUR-Lex Europa. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF

[14] EC. (2003, December) Directive 2003/98/EC (Public sector information (PSI) Directive), Directive on the re-use of public sector information. EUR-Lex Europa. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF

[15] EC. (2004, April) Directive 2004/52/EC (European electronic toll service [EETS] directive), Directive 2004/52/EC on the interoperability of electronic road toll systems in the community. EUR-Lex Europa. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:166:0124:0143:EN:PDF

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

[16] EC. (2007, April) Directive 2007/2/EC (INSPIRE directive), Directive 2007/2/EC establishing an infrastructure for spatial information in the european community (INSPIRE). EUR-Lex Europa. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF
[17] EC. (2010, August) Directive 2010/40/EU (ITS directive), Directive 2010/40/EU on the framework for the deployment of intelligent transport systems in the field of road transport and for interfaces with other modes of transport. EUR-Lex Europa. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0001:0013:EN:PDF
[18] EC. (2013, June) Directive 2013/37/EU, Directive 2013/37/EU amending directive 2003/98/EC on the re-use of public sector information. EUR-Lex Europa. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF
[19] J. Eliasson. (2012, November) How to solve traffic jams? TedxHelvetia. [Online]. Available: http://www.ted.com/talks/jonas_eliasson_how_to_solve_traffic_jams.html
[20] Federal Trade Commission (FTC). (1998, June) Privacy Online: A Report to Congress. Federal Trade Commission Website. [Online]. Available: http://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf
[21] S. Fischer-Hübner and N. Hopper, Eds., Privacy enhancing technologies, Proceedings of 11th International Symposium, PETS 2011, Waterloo, 2011.
[22] S. Fischer-Hübner and M. Wright, Privacy enhancing technologies, Proceedings of 12th International Symposium, PETS 2012, Vigo, 2012.
[23] S. D. Gosling, P. J. Rentfrow and B. S. Williams, A very brief measure of the big-five personality domains, Journal of Research in Personality, vol. 37, no. 6, pp. 504-528, 2003.
[24] A. Gurung, Empirical investigation of the relationship of privacy, security and trust with behavioral intention to transact in E-commerce, Ph.D. dissertation, University of Texas at Arlington, Arlington, TX, USA, 2006.
[25] S. L. Jarvenpaa and N. Tractinsky, Consumer trust in an Internet store: A cross-cultural validation, Journal of Computer Mediated Communication, vol. 5, no. 2, pp. 1-36, 1999.
[26] N. Jentzsch, S. Preibusch and A. Harasser, Study on monetizing privacy. An economic model for pricing personal information. European Network and Security Agency, ENISA, Heraclion, Greece, Technical Report 2012-02-27, 2012.
[27] P. Jeselon and A. Fineberg, (2011, June) The privacy by design privacy assessment framework. Privacy by Design. [Online]. Available: http://privacybydesign.ca/content/uploads/2011/06/2011-06-07-PbD-PIA.pdf
[28] A. Junglas, N. A. Johnson, and C. Spitzmüller, Personality traits and concern for privacy: An empirical study in the context of location-based services, European Journal of Information Systems, vol. 17, no. 4, pp. 387-402, 2008.
[29] D. J. Kim, D. L. Ferrin, and H. R. Rao, A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents, Decision Support Systems, vol. 44, no. 2, pp. 544-564, 2008.
[30] M. L. Korzaan and K. T. Boswell, The influence of personality traits and information privacy concerns on behavioral intentions, Journal of Computer Information Systems, vol. 48, no. 4, pp. 15-24, 2008.
[31] S. Kulk and B. van Loenen, Brave new open data world? International Journal of Spatial Data Infrastructures Research, vol. 7, pp. 196-206, 2012.
[32] R. S. Laufer and M. Wolfe, Privacy as a concept and a social issue: A multidimensional developmental theory, Journal of Social Issues, vol. 33, no. 3, pp. 22-42, 1977.
[33] Y. Li, Empirical studies on online information privacy concerns: Literature review and integrative framework, Communications of the Association for Information Systems, vol. 28, no. 28, pp. 453-496, 2011.
[34] Z. Liu, R. Bonazzi, B. Fritscher, and Y. Pigneur, Privacy-friendly business models for location-based mobile services, Journal of Theoretical and Applied Electronic Commerce Research, vol. 6, no. 2, pp. 90-107, 2011.
[35] N. K. Malhotra, S. S. Kim and J. Agarwal, Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model, Information Systems Research, vol. 15, no. 4, pp. 336-355, 2004.
[36] J. Markkula, Dynamic personal data—New opportunity and challenge introduced by the location aware wireless network, Cluster Computing, vol. 4, no. 4, pp. 369-377, 2001.
[37] Ministry of Transport and Communications of Finland, Oikeudenmukaista ja älykästä liikennettä. Publications of the Ministry of Transport and Communications, Finland, Technical Report 40/2013, 2013.
[38] Ministry of Transport and Communications. (2013, December) Working group chaired by Mr Jorma Ollila: Towards kilometre-based taxation. Ministry of Transport and Communications. [Online]. Available: http://www.lvm.fi/pressreleases/4374272
[39] J. F. Moore, The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems. New York: Harper Business, 1996.
[40] NRI. (2013, February) LKW-MAUT electronic toll collection system for, Germany. Net Resources International Website. [Online]. Available: http://archive.today/fkyZd
[41] NXP. (2010, February) NXP and IBM announce results of landmark road pricing trial. NXP Website. [Online]. Available: http://www.nxp.com/news/press-releases/2010/02/nxp-and-ibm-announce-results-of-landmark-road-pricing-trial.html
[42] Organisation for Economic Co-operation and Development (OECD). (1980, September) OECD Guidelines on the protection of privacy and transborder flows of personal data. OECD Website. [Online]. Available: http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300007

[43] N. Olivero and P. Lunt, Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control, Journal of Economic Psychology, vol. 25, no. 2, pp. 243-262, 2004.
[44] J. Phelps, G. Novak and E. Ferrell, Privacy concerns and consumer willingness to provide personal information, Journal of Public Policy & Marketing, vol. 19, no. 1, pp. 27-41, 2000.
[45] K. Premazzi, S. Castaldo, M. Grosso, P. Raman, S. Brudvig, and C. F. Hofacker, Customer information sharing with E-Vendors: The roles of incentives and Trust, International Journal of Electronic Commerce, vol. 14, no. 3, pp. 63-91, 2010.
[46] D.-H. Shin, The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption, Interacting with Computers, vol. 22, no. 5, pp. 428-438, 2010.
[47] H. J. Smith, J. S. Milberg and J. S. Burke, Information privacy: Measuring individuals' concerns about organizational practices, MIS Quarterly, vol. 20, no. 2, pp. 167-196, 1996.
[48] J. Y. Son and S. S. Kim, Internet users' information privacy-protective responses: A taxonomy and a nomological model, MIS Quarterly, vol. 32, no. 3, pp. 503-529, 2008.
[49] S. Spiekermann, J. Grossklags and B. Berendt, E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior, in Proceedings of 3rd ACM Conference on Electronic Commerce, Tampa, FL, 2001, pp. 38-47.
[50] K. A. Stewart and A. H. Segars, An empirical examination of the concern for information privacy instrument, Information Systems Research, vol. 13, no. 1, pp. 36-49, 2002.
[51] USA. (2009, December) Open government directive. The White House. [Online]. Available: http://www.whitehouse.gov/open/documents/open-government-directive
[52] A. van de Ven and P. Ring, Relying on trust in cooperative inter-organizational relationships, in Handbook of Trust Research (R. Bachman and A. Zaheer, Eds.). Northampton, MA: Edward Elgar Publishing, 2006, pp. 144-164.
[53] B. van Wee, The new dutch per-kilometre driving tax, CESifo, Munich, Germany, CESifo DICE Technical Report 2/2010, 2010.
[54] L. Willenborg and T. de Waal, Statistical Disclosure Control in Practice. Lecture Notes in Statistics, New York: Springer-Verlag, 1996.
[55] L. Willenborg and T. de Waal, Elements of Statistical Disclosure Control. Lecture Notes in Statistics, New York: Springer-Verlag, 2001.
[56] A. Zuiderwijk, N. Helbig, J. R. Gil-García, and M. Janssen, Guest editors' introduction. Innovation through open data: A review of the state-of-the-art and an emerging research agenda, Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no. 2, pp. I-XIII, 2014.