View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Springer - Publisher Connector

Opportunities and Risks Associated with Collecting and Making Usable Additional Data

24

Kai Rannenberg

Abstract

Cars have for a long time been a symbol for the freedom and autonomy of their users. Now autonomous driving raises the question how the data flows related to autonomous driving influence the privacy of these cars' users. Therefore this chapter discusses five guiding questions on autonomous driving, data flows, and the privacy impact of vehicles interacting with other entities: (1) Which "new" or additional data are being collected and processed due to autonomous driving and which consequences result from those "new" or additional data being collected and processed? (Sect. 24.2); (2) Are certain types of data special and do they cause special hindrances? (Sect. 24.3); (3) What is required from the perspective of privacy? (Sect. 24.4); (4) When building architectures, what needs to be kept in mind to avoid creating difficult or even unsolvable privacy problems? (Sect. 24.5); (5) What needs to be considered in the long term? (Sect. 24.6). The questions will be discussed relating as much as possible to the case studies that were introduced at the beginning of this book. Sect. 24.7 concludes this text including an analysis whether more autonomy of driving vehicles leads to more privacy problems.

Thanks go to Tim Schiller, Jetzabel Serna-Olvera, and Markus Tschersich for helpful hints and comments.

K. Rannenberg (✉)
Deutsche Telekom Chair of Mobile Business and Multilateral Security,
Goethe Universität Frankfurt, 60629 Frankfurt am Main, Germany
e-mail: Kai.Rannenberg@m-chair.de

© The Author(s) 2016

M. Maurer et al. (eds.), *Autonomous Driving*,
DOI 10.1007/978-3-662-48847-8_24

24.1 Introduction: Cars, Freedom, and Privacy

Cars have for a long time been a symbol for the freedom and autonomy of their users, be it drivers or passengers: Car drivers can decide on their own, where to drive, which route to choose and often even how fast to travel (or at least when to take a break), and they don't need to report this to anybody. Many pieces of art reflect the opportunity for freedom and escape from (often undue) control that cars offer to their users. Some of the most impressive examples may be episodes 3, 4, 6, and 7 of the 1947 movie "In those days" [1], which describe the more or less successful car journeys of several people oppressed by the German Nazi regime between 1933 and 1945; other examples can be mentioned also (Chap. 3). At the same time, a car offers its driver and holder a protected sphere of privacy: People from outside usually don't hear what is communicated in a car and they cannot easily take a seat and join the conversation. "My car is my castle" is not as popular as "my home is my castle"; still many people see their car as the extension of their home; correspondingly many household goods and activities can be viewed in a car ([2], paragraph 2).

If one takes this perspective, autonomous cars could be just an extension of the traditional concepts of freedom, autonomy, and privacy for their drivers and users. However, "autonomous driving" in first instance makes the driving more autonomous from the drivers. At the same time, it relies much more on interaction with the outside world than a human-driven car. Autonomous cars sense their environment and often even communicate with that environment, e.g. cars nearby. Beyond that exchange with near-by entities, there are plans to control and synchronize cars with traffic centers to optimize their behavior, e.g. their choice of a route. Like any other centralized entity collecting data, this raises privacy concerns, and motivates an analysis of the data flows and the corresponding privacy impact. The situation is more critical if one considers that cars not only can collect a lot of data on their users and the environment, but also store them for a long time and then communicate them to other entities.

Therefore this chapter follows five guiding questions on autonomous driving, data flows, and the privacy impact of vehicles interacting with other entities:

1. Which "new" or additional data are being collected and processed due to autonomous driving, and which consequences result from those "new" or additional data being collected and processed? (Sect. 24.2)
2. Are certain types of data special and do they cause special hindrances? (Sect. 24.3)
3. What is required from the perspective of privacy? (Sect. 24.4)
4. When building architectures, what needs to be kept in mind to avoid creating difficult or even unsolvable privacy problems? (Sect. 24.5)
5. What needs to be considered in the long term? (Sect. 24.6).

The questions will be discussed relating as much as possible to the use cases that were introduced at the beginning of this book. Section 24.7 concludes this chapter.

24.2 Additional Data Collected and Processed due to Autonomous Driving

To assess the opportunities and risks associated with collecting additional data and making it usable, it is useful to first try to identify those data. This will be done along the four use cases defined in Chap. 2, but first a short overview on data that is being collected or may be collected in a non-autonomous car will be given.

24.2.1 Personal Data Collected and Potentially Transmitted in Today's Networked Cars

While the analysis in this chapter will concentrate on "new" or additional data, it should be mentioned here, that quite a few sensitive personal data are already collected in today's cars and sometimes even transmitted. Some examples are

- All types of location data and navigational data: Typical data are destination, travel time, travel habits ("every weekend to Stuttgart") and preferences for routing (scenic vs. fast vs. ecologically friendly vs. on the edge of legality). Especially when a car is tracked by a dispatching system, a theft control system, a car insurance system, or a road pricing system, a lot of information is collected about its whereabouts and in many cases transferred to the related central entities. Some of these systems store data in decentralized fashion due to their sensitivity, but others don't. An example that became quite prominent recently is the new European eCall system [3–5]: eCall is activated automatically as soon as in-vehicle sensors detect a serious crash. Once set off, the system dials the European emergency number 112, establishes a telephone link to an emergency call center and sends details of the accident to the rescue services, including the time of incident, the accurate position of the crashed vehicle and the direction of travel (most important on motorways and in tunnels). An eCall can also be triggered manually by pushing a button in the car, for example by a witness to a serious accident.
- Data on driving dynamics: This type of data, for instance on acceleration, gives information on the behavior of the car but also on the behavior of the driver, such as the driving style (e.g. calm vs. aggressive vs. fast vs. on the edge of legality).
- Data on driving behavior: These data can be derived from location data over time. For example, comparing the location of a car on a highway with the location 15 min before may tell about the average speed of the car and whether a speed limit was violated or possibly violated.
- The environment: The car may be collecting data from the environment to document the ride or specific traffic situations in case such documentation would be considered helpful later. Examples include dashboard cameras to document and maybe transmit

what is happening in front of the car. Data from the environment may well be other people's personal data, e.g. number plates of other vehicles or faces of people.

This rough overview also raises the question: What kinds of data are actually personal? Some of the data listed may not seem to be "personal". However, experience over years of initiatives aimed at privacy protection has shown that there are no guarantees that data cannot be related to persons and cannot be misused. One consequence of this lesson is that "personally identifiable information (PII)" is nowadays not only the information that directly identifies a person, but "any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal" ([8], clause 2.9).[1] The PII principal ([8], clause 2.11) is then the individual whose data are being processed. In our cases, PII principals may include drivers, passengers, or car owners, but also passersby who are sensed and can be identified in any way.

Still, the fact remains that the practical sensitivity of data at a certain point in time depends a lot on context, e.g. the location information of a car may be more sensitive if the car is parked near a red-light district. More examples will be seen in the discussion of the use cases of autonomous driving and the related interests of the parties involved. Moreover, analyzing the cases helps to illustrate new situations and the respective issues.

24.2.2 Personal Data Collected in Autonomous Cars

This section discusses the data collection in autonomous cars along the four cases as introduced in the beginning of this book (Chap. 2).

24.2.2.1 Use Case 1: Interstate Pilot Using Driver for Extended Availability

The driving robot takes over the driving task, but only on interstates or interstate-like expressways. During autonomous journeys, drivers become passengers who can take their hands off the steering wheel and feet off the pedals and pursue other activities. The driving robot coordinates a safe handover to the driver and may even stop the car at a safe place if needed.

[1]The relation between data and information is too subtle and too complex to be sufficiently explained within the limits of this chapter. All the same, considering data and information as roughly equivalent should suffice for the purpose of this chapter. Using only one term would put the paper at odds with some of the referenced literature.

The new and additional data the car can collect and learn in this case are data on:

- Driver capabilities, e.g. whether the driver is able to take back control from the robot or not, and how much time such a take-back-operation takes: Both types of data may be of interest as up-to-date data for immediate reaction but also as a basis for longitudinal evaluations.
- Driving behavior: On top of other data on driving behavior that are already available nowadays, this case enables the collection of additional data, i.e. under which circumstances the driver delegates control and/or demands it back.
- The environment: Additional data from the environment can be collected to pursue the autonomous ride. Also, documenting the ride or specific traffic situations may be considered helpful to handle potential conflicts. As in Sect. 24.2.1, data from the environment may well be the personal data of other people, e.g. number plates of vehicles or people's faces. So environmental data contain a mix of personal data of several people, which makes them especially delicate.

Following the discussion on the legal and liability implications of autonomous driving (see the part of this book on law and liability (Part V) one can indeed assume that there will be an interest in collecting data to document potential accidents and investigate what behavior of the car, the robot, the driver, or other parties may have caused the accidents. This would accord with other cases of law enforcement agencies having a large "appetite" for data that may become available through the computerization of an activity, as computerized activities are usually easy to log in computerized logs.

24.2.2.2 Use Case 2: Autonomous Valet Parking

The driving robot parks the vehicle at a nearby or remote location after the users have exited and cargo has been unloaded. The driving robot drives the vehicle from the parking location to a desired destination. The driving robot re-parks the vehicle. The driver saves the time of finding a parking spot as well as of walking to/from a remote parking spot. In addition, access to the vehicle is eased (spatially and temporally). Additional parking space is used more efficiently and search for parking is arranged more efficiently.

The new and additional data the car can collect and learn in this case are data on:

- The duration of a stay: How much time do users spend at their destination?
- The area of interest: Where do the users spend more or less time?
- Times of travel and duration of gaps in between: When do users spend more or less time leaving the car alone (e.g. always at Saturday night there is a ride to somewhere and then a long break of more than 8 h)?
- Under which circumstances is the car left alone?

- Visiting habits: How often does the user go where? For example, "every weekend to a certain supermarket, bar, or discotheque."
- The environment: These data may in principle be the same as in Use Case 1, but depending on the environment, there may be different data. Driving through a parking lot the car may "catch" more number plates than on a highway, but fewer people (and faces) in the cars. However, on the way to the parking lot there may be more pedestrians, e.g. crossing the streets, so more faces to recognize.

As there is no direct interaction between the driver and the car, no data on driving behavior are collected.

24.2.2.3 Use Case 3: Full Automation Using Driver for Extended Availability

Use Case 3 is similar to Use Case 1, as in both cases the driving robot performs the driving task with drivers just being passengers who can take their hands off the steering wheel and their feet off the pedals and pursue other activities. However, in Use Case 3 the driver can delegate the driving task to the driving robot in many permitted areas, not just expressways. So the new and additional data the car can collect and learn are basically the same as in Use Case 1, but there may be more options for the driver to delegate and take back control. This can lead to more data being available on driver behavior, especially in circumstances where the driver delegates control and/or takes it back. Similarly to Use Case 2, the data on the environment may be richer and more sensitive than the data collected on a highway described in Use Case 1.

24.2.2.4 Use Case 4: Vehicle on Demand

The driving robot drives the vehicle autonomously in all scenarios with occupants and/or cargo, but also completely without any payload. The driving robot makes the vehicle available at any requested location. Passengers use the travel time completely independently for activities other than performing the driving task. The cabin design is completely free of any requirements for any driver workplace whatsoever, but it may have a camera directed at the passenger space.

While this case is the most demanding from the perspective of autonomous driving, there may be less additional data collected than in Use Case 3. In particular, there is no additional data collected on driving behavior, as there is no driver in the loop anymore. The additional data collected are:

- Travel behavior (e.g., when do passengers want to take breaks?)
- General behavior (or misbehavior) of all passengers in the car
- Data collected on the environment, e.g. to document an accident and what may have caused it (if data on passengers are considered useful for accident documentation).

24.2.3 Consequences for Control Over Data and Misuse Resulting from Data Storage

In principle, the storage of any kind of data opens possibilities for any kind of processing that would not exist without that storage. While this seems to be a theoretical triviality, the practical consequences of storing data are that they can be used and misused later, maybe under circumstances of which the user was not originally aware. This implies a longer-term responsibility for these data. The responsibility has to lie with the body that can control the data and make decisions about their usage.

If one can assume that the data stored in a car are under the sole control of the car's owner or driver, then determining responsibility for these data may be relatively easy. Otherwise, the responsibility for storage and any kind of misuse would expand to the body that controls the storage or transfer or both. This only applies, however, if the data can leave the car and the domain of its owner without the owner being in control of this data transfer.

There are at least two indications that powerful bodies will ask for data stored in a car to be transmitted out of the car:

1. Law enforcement agencies very often take the approach that data stored for any technical or commercial purpose should also be made available for law enforcement purposes. Lawmakers have often followed this position. The example closest to cars and location data is that of mobile phone communication. From the beginning of the 1990s, the GSM standard for cell-based mobile communication has been established and the location information of subscribers processed in the networks. Rules were soon established to enable law enforcement agencies to access all types of data in the GSM networks, including location data: an example is the German Fernmeldeüberwachungsverordnung [6], which was established as early as 1995.
2. Internet enterprises such as Google are inspired and driven by connectivity and transmitting data. An example is a statement of Jared Cohen, Director of Google Ideas and Eric Schmidt, Executive Chairman of Google in the conclusions of their joint book "The New Digital Age" [7], p. 254: "Attempts to contain the spread of connectivity or curtail people's access will always fail over a long enough period of time–information, like water, will always find its way through."

Not everything that these powerful organizations have been asking for has happened, but the examples give an impression of the challenges accompanying data storage, even in a contained fashion.

24.2.4 Consequences from Data Transfer to Third Parties

Data being transferred to entities outside of the domain of the car owner or driver (third parties) can enable those entities to pursue their interests. These interests may or may not

conform to the interests of the parties identified by the data (also called data subjects), typically the driver or the owner of the car. This section will give examples of the following third parties: vehicle manufacturers, insurance services, fleet operators, government-authorized parties, peer ad-hoc networks, e.g. other traffic entities or other autonomous vehicles, and traffic centers. This sequence of sections follows the rising complexity in the setting of the third party entities.

24.2.4.1 Vehicle Manufacturers

Vehicle manufacturers may be interested in documenting vehicle behavior, e.g. to learn about the vehicle's behavior in extreme situations and about the quality of their (often very complex) software, and to improve the systems. These data are similar to the kind of data that manufacturers and operators of telecommunication systems collect for quality assurance and maintenance purposes. At the same time, these data also deliver sensitive information about the driver, e.g. the typical driving speed and the number of emergency brakes or missed handovers from the driving robot in Use Cases 1 and 3.

24.2.4.2 Insurance Services

Insurance services are often interested in more information about their customers to assess the level of risk associated with them or gain other customer insight. Depending on the type of insurance, different information can be of interest, e.g. the insurance risk for an accident can be derived from driving behavior (risk averse or less risk averse driving style), and from location information for theft insurance (regions with more or less theft risk for the specific vehicle). All cases offer rich data here, Use Cases 1 and 3 more on driver behavior, all cases on location information, Use Case 4 also on occupants' behavior and emergency calls. These assessments may be fairer to insurance customers, as they award cost-reducing behavior, but they put users under more surveillance without a clear description of the related risks and opportunities. Often insurance services make decisions based on scoring systems or details unknown to customers, as these details are considered "trade secrets" the insurance companies wish kept confidential to protect themselves in a competitive market. Customers may then be surprised about decisions, e.g. the denial of an upgraded contract or a fee raise.

24.2.4.3 Fleet Operators

Fleet operators such as rental car companies are in a situation similar to that of insurance companies. To raise their commercial success, they try to assess the risk associated with handing out a car to a certain customer and to consider the results of their assessments for their pricing. Therefore the consequences for customers are also similar to those in the case of insurances, e.g. with regard to (non)-extensions of contracts or fee rises. Also for this scenario, all use cases offer data. A major difference to the case of insurance companies is the fact that fleet operators usually own the cars, so they have more control over the cars than an insurance company has over an insured car. This difference is important for any concept of a "private data vault" to store sensitive data of rental customers or

drivers (see Sect. 24.5). In the fleet-operator scenario, such a "private data vault" would either need to be specially installed within the car to protect it from access by the fleet operator or it would need to be brought along by the rental customer or driver.

24.2.4.4 Commercial Location-Based Services

Advertisers are interested in directing the right messages to their respective target groups. This may include placing advertisements at the right location, e.g. special offers of shops near the next exit, encouraging commuters in a traffic jam to leave the highway and go shopping. Also, travelers in a traffic jam on the way to a major airport can be targeted with offers from a smaller regional airport (as seen, for instance, on the highway north towards San Francisco airport, where flights starting from San Jose have been advertised). Advertisers are thus interested in traffic flows (and jams). Moreover, they always like to know more details about their target groups, so any kind of behavior that allows conclusions, e.g. on the type of traveler (business, commuting, leisure) will be welcome.

24.2.4.5 Government-Authorized Parties

Government-authorized parties such as police forces or intelligence agencies can use the data for surveillance to detect behavior they want to sanction or prevent. In the case of the traffic police, this can be any kind of behavior deemed unsafe or in violation of traffic laws, e.g. difficulties or strange behavior in interaction with the driving robot. Polices forces investigating crimes or aiming to prevent crimes as well as intelligence agencies may be interested in analyzing navigation and movement data to learn about the social networks of travelers, e.g. who may meet whom where. There is also the strong potential that the interested government and intelligence agencies have their very own interpretation of what they are authorized to do beyond the assurances and guarantees of privacy laws and privacy protection. This may especially hold for data on the environment the car would be collecting. Having many or all cars collecting data from the environment can be considered a specific form of crowd-sourcing. Some municipalities are considering crowd-sourcing to collect pollution data. This may collect fewer or no personal data from the environment, but conceptually it is not too far from a car spying on its environment.

24.2.4.6 Peer Ad-Hoc Networks

Peer ad-hoc networks (e.g. other traffic entities or other autonomous vehicles) can be interested in any kind of data used or analyzed by a specific car to optimize path tracking and stabilization or which results from the optimization process. This data may help the peer ad-hoc networks to assess the road conditions other vehicles are experiencing especially at locations that are touched by their own routes. If the data are anonymized and stay with the involved peers, the consequences are less severe than data transfer to (central) entities aggregating data like the other entities discussed in this chapter.

24.2.4.7 Traffic Control Centers

Traffic control centers' interests depend a lot on the interests of their operators and owners. Control centers aiming for efficient traffic flows, and to mitigate the effects of accidents on traffic flows, are interested in any kind of data that help them to assess current and future traffic situations: Driving conditions can be derived from environmental information or from the assessments of driving behavior as delivered in all cases; potential congestion can be derived from travel plans and navigational data. The other aims of these centers may include collaborations with other entities to refinance their costs or even deliver a profit to their owners or operators. This is helped by the fact that the other entities discussed earlier in this section can make use of the data collected by traffic centers.

The degree to which traffic control centers would be interested in collaborating with other entities interested in their data, and offer some reimbursement, may well depend on their status and financial situation. A private for-profit control center would need to find funding; a public traffic center may be under less pressure here. However, for many recent major investments in public infrastructures, there were aims to operate them as public-private partnerships to mitigate the lack of public money for investments. This holds for toll-fee collection and was also the plan (though not a successful one) for the Galileo satellite network. Also, public broadcasting companies are becoming more and more dependent on private co-funding, e.g. from advertisements.

24.3 Are Certain Types of Data Special and Do They Create Particular Obstacles?

It is extremely difficult to predict the potential use of data for legitimate or illegitimate purposes, and it has proven impossible to guarantee that no kind of data would be used or misused, even in the long run. One reason is that, with today's connectivity, combining data is easy. Data on the agility a driver shows when taking back control from a driving robot may look harmless, but if put in relation to the same data 10 years before or hence, they can give the impression of driving abilities tending to rise or fall. This may put the driver at an unfair disadvantage, e.g. when insurance fees are calculated. Similar scoring activities by credit rating agencies have shown to be often very wrong with regard to an individual, even if they may have a statistical value. Therefore there are no explicit rules to consider certain data special and have special hindrances for their usage. One can get the feeling that data allowing conclusions on to be drawn people's health and/or (political) views are especially sensitive, but there are no clear indications that these are always more sensitive than data on their financial situation, for example.

The legal consequence of this difficulty is the principle of asking for the processing of each and every piece of data to be authorized, instead of giving general clearances (see also the descriptions of "Purpose legitimacy and specification" and "Collection limitation" in Sect. 24.4.1). All data thus needs to be checked: Are they absolutely necessary to

provide the service for which they have been collected? Was that type of processing appropriate?

24.4 Requirements from the Perspective of Privacy

This section discusses the requirements from a privacy perspective, starting with an introduction of internationally established principles and their relation to the use cases (Sect. 24.4.1). Then additional surveillance measures for a "data-protected" usage of the additional data are discussed in Sect. 24.4.2, before Sect. 24.4.3 focuses on limiting access rights and on encryption.

24.4.1 Principles

For any personal data collected and transmitted beyond the domain of the data subject, there must be a clear justification with regard to relevant privacy principles and related requirements. Privacy principles and requirements depend on the respective national, regional and sometimes sector-specific legislation, so a complete analysis would be impossible. Fortunately there is now the international standard ISO/IEC 29100 Privacy Framework that was completed in 2011 and lists eleven privacy principles [8]. These privacy principles were derived from existing principles developed by states, countries and international organizations, e.g. the OECD and the EU. The editors came from Germany and the USA, and experts from many countries participated intensively in the development. One focus of the ISO/IEC 29100 Privacy Framework is the implementation of the privacy principles in ICT (Information and Communications Technology) systems; another is on developing privacy management systems within organizations' ICT systems. The privacy principles aim at guiding the design, development, and implementation of privacy policies and privacy controls. A sketch of related requirements can also be found in a recent recommendation of the very influential "Deutscher Verkehrsgerichtstag," an annual conference of legal experts focused on traffic regulation [9]. The eleven principles are:

1. Consent and choice
2. Purpose legitimacy and specification
3. Collection limitation
4. Data minimization
5. Use, retention and disclosure limitation
6. Accuracy and quality
7. Openness, transparency and notice
8. Individual participation and access

9. Accountability
10. Information security
11. Privacy compliance

This section will concentrate on explaining the principles considered most important and give some examples from the use cases[2]:

- Consent and choice: The Consent principle was introduced over time to ensure that PII (personally identifiable information) principals can control whether or not their PII is being processed except where applicable law specifically allows the processing of PII without consent. It is explicitly mentioned that consent needs to be informed consent, so PII principals are to be informed about what they agree to, and it also needs to be opt-in consent. It turned out that demanding appropriate choice became important to avoid users giving de facto consent, as they have no alternative to get the respective service. In Use Cases 1, 2 and 3, consent will be needed from the owner, the driver and any identified passenger. In Use Case 4, passengers' and, if applicable, drivers' consent is required. The most critical question, however, arises around consent for scanned environmental data. For example, private observation cameras are usually not allowed when they cover public space and can collect data from people there. For data from public observation cameras there are strict rulings that follow the following principles.
- Purpose legitimacy and specification: Adhering to this principle means: ensuring that any purposes comply with applicable law and rely on a permissible legal basis; communicating any purpose to the PII principal before the information is collected or used for a new purpose; using language for this specification which is clear and appropriately adapted to the circumstances; and, if applicable, giving sufficient explanations for the need to process sensitive PII. A purpose can require a legal basis or a specific authorization by a data protection authority or a government authority. If the purposes for processing PII do not conform to applicable law, processing should not take place. For all use cases this means especially that the purposes need to be specified explicitly and in a clear way. This will be a special challenge for the scanning of environmental data.
- Collection limitation: The collection of PII is to be limited within the bounds of applicable law and those data that are strictly necessary for the specified purpose(s). In our use cases, this applies especially to any data on the behavior of any driver and identified passenger. If the purpose is autonomous driving, any data collection will need to be justified in relation to autonomous driving (and not any other use, even if, for example, it seems commercially attractive).
- Data minimization: Data minimization is closely linked to collection limitation, but refers to strictly minimizing the *processing* of PII. Data processing procedures and ICT systems are to minimize the PII processed and access to it. Default options should,

[2]ISO/IEC 29100 has more extensive explanations of the principles.

wherever possible, not involve the identification of PII principals, reduce the observability of their behavior, and limit the linkability of the PII collected with other PII (and thereby also the traceability of the PII principal). Moreover, one should delete and dispose of PII whenever the purpose for PII processing has expired, there are no legal requirements to keep the PII, or whenever it is practical to do so. For all four use cases, this principle limits the transfer of any data to any central entities, such as traffic control centers: PII that is only needed to manage the situation in and around the vehicle shall not leave the vehicle without the PII principal's permission. Data minimization also demands the storage of sensed data to be limited, especially when the data can easily be recollected when needed again. Last but not least, limiting the linkability of the PII collected calls for anonymization and aggregation of any data that is not needed for individual cases.

- Information security: Information security refers to protecting PII with appropriate controls at operational, functional and strategic levels to ensure the integrity, confidentiality and availability of the PII, and protect it against risks such as unauthorized access, destruction, use, modification, disclosure or loss throughout the whole of its life cycle. Information security spans a wide spectrum from choosing an appropriate PII processor to limiting access to PII to those individuals who require such access to perform their duties. Sects. 24.4.2 and 24.4.3 describe related measures.

Data usage beyond that absolutely needed to provide the service for which the data was collected requires explicit consent. So for any PII collected and transferred beyond the domain of the PII principal, there must be a clear and convincing rationale with regard to relevant privacy principles. The rationale must be convincing for the PII principal in terms of what is gained and what is given up. The rationale must also be convincing for the regulator, who will check whether the PII principal is being misled, e.g. by stating a data processing necessity that does not exist when, following the data minimization principle, an alternative methodology or technology could be chosen. The regulator will also check, whether fundamental rights would be endangered by processing the data; fundamental rights cannot simply be given up by users through consent, as they may not understand the consequences. A related example would be asking users to store and process their voting behavior.

Any PII that may enable users to be discriminated against according to their beliefs, thoughts or actions (e.g. topics of interest and related locations and destinations (e.g. towards a political demonstration), also in relation to other people's locations and destinations of) is in many practical cases especially critical.

One example for a reasoning can be found in the landmark decision of the German Constitutional Court from 1983 [10], that established the fundamental right of "Informational Self-Determination" in Germany and asks to beware of a "chilling effect" on citizens' participation in democratic processes: A person who is uncertain as to whether unusual behavior is being taken note of, used, or transferred to others will attempt to avoid standing out through such behavior. Persons who assume, for example, that attendance of

an assembly or participation in a citizens' interest group will be officially recorded, and that this could expose them to risks, will possibly waive the exercising of their corresponding fundamental rights. This would not only restrict the possibilities for personal development of those individuals but also be detrimental to the public good, as self-determination is an elementary prerequisite for the functioning of a free democratic society based on the freedom of action and participation of its citizens.

Similar considerations are especially relevant in countries with unstable political governance, where citizens have to fear that a future government may not tolerate behavior that would currently be perfectly legal. This may include travelling to a political meeting.

Moreover, the Snowden revelations [11] have shown that there are severe weaknesses to be considered in the data security governance of many entities storing PII, especially data attractive to intelligence agencies. These developments can be expected to be included in future risk analyses and considerations.

24.4.2 Additional Surveillance Measures for "Data-Protected" Usage of Additional Data

Additional surveillance measures for a "data-protected" usage of the additional data can be foreseen. They are motivated, for example, by the ISO/IEC 29100 principle of accountability [8].

The accountability principle means that the processing of PII entails a duty of care and the adoption of concrete and practical measures for its protection. This will apply to any party processing PII. The measures are supposed to not only secure the proper processing, but also enable and ease supervision by regulatory authorities, e.g. data protection commissioners.

The information security principle (cf. also Sect. 24.4.1) calls, for example, for controls at operational, functional and strategic levels to ensure the integrity, confidentiality and availability of the PII, and protect it against risks such as unauthorized access, destruction, use, modification, disclosure or loss throughout the whole of its life cycle.

Typical issues of importance are to audit who had or has access to the PII and who worked or works with it in which way. Additional surveillance measures will therefore apply to any additional entity that may have access to the PII. Experience with auditing has shown that it can lead to additional privacy problems, as audit records on processing may be used in an even more discriminatory way than the data themselves. An example could be an entry in a traffic control center's audit log that PII on the reaction times of a certain driver's interaction with the driving robot were examined by a task force to analyze driving behavior.

Moreover, additional surveillance measures should not lead to oversurveillance of the individuals working with the system, at least in regions where privacy in the workplace is

protected. So a fine balance has to be found depending on the relation between customer protection and employee protection.

24.4.3 Limiting Access Rights and Encryption

Limiting access rights and encryption are typical instruments of information security. Limiting access rights is also mentioned under the "Information Security" principle of ISO/IEC 29100. It follows the concept of "need-to-know", limiting access to PII to those individuals who require such access to perform their duties, and limiting the access of those individuals to only that PII which they require to perform their duties. Access rights can be defined by defining exactly which entity can access which PII. This asks for a fine-grained specification of the system, and can best be achieved if privacy is already considered during the design phase, e.g. when designing which data are collected by the vehicle and for which application they are needed.

Another way to limit access rights is to define that only groups of entities can jointly access certain data, e.g. any kind of audit records. This four-eye principle (or n-eye principle) helps against unauthorized use of data and can especially apply to audit records of system behavior involving PII. One could specify, for example, that these kinds of data are only made available to address a defined system failure, and that both the PII principal and the interested party, e.g. an authorized repair shop, need to agree on the access. The n-eye principle can also be implemented by encryption if parts of keys are distributed among the respective stakeholders.

Encryption is not directly mentioned in ISO/IEC 29100 as its use is sometimes considered controversial in some ISO/IEC member states. However, encryption is mentioned as an example of a requirement for transmitting medical PII over a public network ([8], Clause 4.4.7). It is also being asked for more and more by Privacy Commissioners, who have understood its advantages, especially for creating virtual vaults or tunnels to protect PII even without the cooperation of the entity storing or transporting the PII. If encryption is used, it is important to define clearly who will be allowed to hold the keys for the respective encryption and decryption. PII providing clues to an individual's behavior and abilities may need to be protected by an asymmetric encryption system, and then by encrypting the PII with the public key of the respective individual. This would ensure that the PII can only be decrypted with the corresponding private key of the individual.

24.5 Architectural Considerations

Any architectural considerations need to consider the interests of systems' stakeholders. PII stakeholders mentioned in this text so far include drivers, passengers and owners of cars. Other stakeholders may be individuals who need to work with the PII, perhaps also

bystanders on the street or other traffic participants if they can be identified by the system.[3] It turns out to be useful to consider non-professional users of systems especially, as they usually have less opportunity to protect themselves [12]. They are usually also those entities that Privacy Commissioners are meant to look after.

In general, architecture characteristics can be derived from the principles discussed in Sect. 24.4.1. The principles of collection limitation, data minimization, and information security are especially relevant for architectural considerations. Any architecture that allows a service to be provided that collects, uses and spreads less PII not only reduces the damaging consequences of any misuse, but also eases the securing of information.

Three architecture characteristics and elements are especially recommendable:

1. Decentralized approaches: If PII is not transferred to central entities, such as traffic centers, the risk of misuse is reduced. Examples include:
   - If, in any of the cases, a situation can be resolved directly between two vehicles, this is better than involving a traffic control center or other external entity. Sometimes the issue of the trustworthiness of the information provided by other vehicles is brought up. A quick solution seems to be to identify the other vehicle individually and to check it against a central registration database, similarly to a police car checking registration plates of cars. This may be a nice sales scenario for selling directory services, but viewing it as a gain for privacy or security is short-sighted. It would transfer an exceptional police activity into a regular activity performed by perhaps every vehicle, and so establish a massive surveillance infrastructure. Moreover, being able to identify a car precisely does not give any guarantee for the information provided by that vehicle. This information may still be manipulated and misleading even if a valid identifier is sent by the vehicle originating the information.
   - The concept of a user-owned "Private Data Vault" (PDV) to store PII should be explored in more detail to enable the storage of sensitive data under the user's control. This PDV could store the PII of the respective individuals and protect it against unwanted access, so that access is not possible without those individuals' consent. Especially for drivers using cars used by several drivers, and for rental customers or drivers, this would be useful. A PDV could be installed within the vehicle (in the special case of vehicles used mainly by a single person) or would ideally be brought by the respective driver when using the car. The PDV should use appropriate hardware protection for storing the data, and can be the initialization of trustworthy data stores. A combination with other personal devices such as mobile phones might be possible in future, but first these devices need to become more secure and better able to protect themselves, especially against outside approaches

[3]This may be a motivation to design the system in such a way that it does not identify bystanders on the street or other traffic participants.

to read their data. Related concepts exist for road toll charging, see e.g. [13], and pay-as-you drive insurance, see e.g. [14].
- If data needs to be stored that is not only personal data of the car user, but also of other parties, such as environmental data (it may identify other people, but also the route the user used), the four-or-n-eye principle should be applied for access control.
- In Use Case 2, traffic control centers or other entities involved in the choice of parking spaces should not ask the drivers or passengers for all kinds of priorities for a parking space or route, but instead give some options, so that the user or, a local system assisting the user, can choose. This reduces the risk of a centralized processing of users' attitudes with regards to prices and locational preferences.

2. Anonymization: Information that needs to be collected for a justified purpose does not necessarily need to be collected in a way that identifies the respective individual. Even information that is collected in a way that identifies individuals may not need to be processed further in such a way. This holds especially for any information that is only needed in an aggregated form:
   - Traffic and congestion analysis does not need to identify individual cars or even drivers.
   - Interaction with peers, e.g. exchanging data for traffic safety with other vehicles, does not require identification (see the discussion above under "Decentralized approaches").
   - Not even access control for cars (e.g. to decide access to parking spaces) needs to identify cars individually. The concepts of Partial Identities (ISO/IEC 24760-1, [15]) and Privacy-friendly Attribute Based Credentials [16] allow the limiting of information presented in such cases to what is really needed to gain access. For example, in Use Case 2 the certified information that a parking space was booked for the autonomous valet-parking of a vehicle does not need to identify the individual vehicle towards the access control system of the parking space. Transferring a presentation token that only identifies the vehicle if it is used twice (and is therefore misused) should suffice.
3. Systematic deletion of PII: Data deletion is often neglected in concepts and life-cycle models for ICT systems. Especially in the case of PII, this can lead to dangerous misuse and consequential liabilities. Therefore, draft architectures in any of the cases should already be including concepts for systematic data deletion; this requires careful consideration as to how long which data need to be kept for which purpose. Within the German Standardization Organization DIN, and also at ISO/IEC, standardization initiatives for data deletion have been started, see e.g. [17]. These initiatives build to a major degree on the data deletion concept of the German Toll Collect road pricing scheme for trucks.

24.6 Long-term Considerations

It seems likely that any infrastructures for autonomous driving will be large, and therefore that any planning for their introduction, use, and maintenance needs to be long-term. A few remarks about long-term experiences should, then, be useful:

1. Application creep: Once a technical infrastructure is established for some applications, new additional applications "piggy-backing" on the same technology and infrastructure but with more privacy risks can be easy to implement. This has been experienced, for instance for the GSM mobile communication network, which has a lot of powerful functionality; or for localization, whose de-facto-introduction and exploitation in some countries has been a grey area. Related fears exist for road tolling systems and their surveillance infrastructures established for trucks or other commercial vehicles only. The extension to private cars may be easy.
2. Creep from test systems to real systems: Experience in Internet software development shows that the step from a test system, or even an experimental prototype with reduced or no security or privacy protection, to a real production system may be as easy as changing the web link on a public portal to point to a new backend system. Such a change may lead to test systems being rushed into real production, while these systems may not be protected like real systems. Particularly projects that are short of resources and need quick success can be tempted by this strategy.
3. Mandatory pseudo-unique identification: More and more computer devices store and issue identifiers that identify these devices more or less uniquely and reliably. One example is the GSM International Mobile Station Equipment Identity (IMEI). In theory, the IMEI is a unique identifier for every GSM mobile communication device; in practice, it can be manipulated. A similar situation exists for the Media Access Control address (MAC address) in Internet networks, which theoretically is a unique identifier assigned to network interfaces. Both identifiers also relate to cars equipped with the respective communication technology. While the security of these systems is low, they make (unofficial) data collections very easy and hence create major privacy problems. Moreover, they foster a recurring "appetite" in interested parties for more identification of users in communication networks or Internet services. This trend needs to be recognized, considered, and overcome [18].

24.7 Concluding Considerations

One may think that a higher degree of autonomy in driving would lead to more data processing to enable the autonomous driving, and consequently to more surveillance. Actually, this is not necessarily the case. The two main factors leading to the collection and spreading of additional data through autonomous driving are:

1. The interaction between the vehicle and the driver(s), passenger(s) and possibly owner (s) becomes more intensive, which leads to the storage and processing of additional data.
2. The interaction of the vehicle with other entities, especially with any kind of traffic control center becomes more intensive, which leads to additional transfer of potentially sensitive data out of the vehicle.

An autonomously driving vehicle that drives sufficiently autonomously, that it does not need to interact with a driver, does not need to collect more data from a driver than any "conventional" car. Also, if the vehicle is able to autonomously navigate through traffic and reach its destination, it would not communicate more data than any other vehicle, and even less data, than a conventional car using a centralized navigation system and that is under surveillance, e.g. by a system that constantly collects the geo-coordinates of the car.

Of course, some of the close-to reality intermediate scenarios, e.g. a vehicle handing over to a driver in critical situations (see e.g. Use Case 1) combined with a centralized surveillance in critical situation can lead to more surveillance and consequently more privacy problems. So while, in theory, vehicles driving more autonomously does not necessarily lead to more privacy problems, there is a realistic threat that in practice this will happen if the design and architecture do not carefully avoid privacy problems.

Therefore, an approach of privacy-by-design for autonomous driving-scenarios is needed. At least for the following questions, one needs to perform a thorough check:

- Is the collection, processing, or transmission of data really needed for a real improvement in the driving situation?
- Is this advantage worth the additional privacy risks?
- In potential dilemmas between more functionality and more safety on the one hand, and less privacy on the other, can the PII stakeholders (often drivers, passengers, owners) be enabled to decide for themselves and in an informed manner?
- Do the data stay under the control of the PII stakeholders, or do they leave their domain of control?

There is clearly a challenge to protect the freedom that has been associated with personal cars for a long time, and that is one of the reasons for their success. Perhaps a unique selling point for the established car industry, and especially premium manufacturers and brands, is to not simply follow the easy trend of Internet businesses in letting information flow everywhere unless they get stopped by legislation or customer outrage, but rather to facilitate proper protection for their customers. The car industry has shown in other areas, for instance in the reduction of energy consumption, that one does not need to accept primitive solutions, but can overcome adverse effects and reduce resource usage by careful planning and engineering. The triggers for this approach will come anyway.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

1. Helmut Käutner, Ernst Schnabel: In Those Days (German: In jenen Tagen); Camera-Filmproduktion 1947; more information on http://en.wikipedia.org/wiki/In_Those_Days; http://de.wikipedia.org/wiki/In_jenen_Tagen_%281947%29; last visited 2014-08-15
2. Angelina Göb: "Nimm Zwei – Carpool Lanes"; 2013-08-06; www.urbanfreak.de/carpool-lanes/, last visited 2014-08-15
3. Telematics News: eCall in German privacy debate; published: 01 February 2012; http://telematicsnews.info/2012/02/01/ecall-in-german-privacy-debate_f3011/; last visited 2014-08-15
4. Jan Philipp Albrecht: eCall - Überwachung aller Autofahrten muss gestoppt werden; www.greens-efa.eu/ecall-11553.html; last visited 2014-08-15
5. European Parliament: Decision of the European Parliament and of the Council on the deployment of the interoperable EU wide eCall service; P7_TA-PROV(2014)0359; http://www.europarl.europa.eu/RegistreWeb
6. German Federal Government: Verordnung über die technische Umsetzung von Überwachungsmaßnahmen des Fernmeldeverkehrs in Fernmeldeanlagen, die für den öffentlichen Verkehr bestimmt sind (18. Mai 1995, BGBl. I S. 722)
7. Eric Schmidt, Jared Cohen: The New Digital Age: Reshaping the Future of People, Nations and Business; Alfred A. Knopf 2013, ISBN-10: 0307957136
8. ISO/IEC 29100:2011 Information technology — Security techniques — Privacy framework, First edition, 2011-12-15, freely available via http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html
9. Deutscher Verkehrsgerichtstag 2014, 29. bis 31. Januar 2014 in Goslar, Arbeitskreis VII: Wem gehören die Fahrzeugdaten? www.deutscher-verkehrsgerichtstag.de/images/pdf/empfehlungen_52_vgt.pdf; last visited 2014-08-29
10. German Constitutional Court: Volkszählungsurteil: Urteil v. 15. Dezember 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83; www.servat.unibe.ch/dfr/bv065001.html and Mitglieder des Bundesverfassungsgerichts (Hrsg.): Entscheidungen des Bundesverfassungsgerichts. 65, Mohr, Tübingen, S. 1–71, ISSN 0433-7646; unofficial English translation on https://freiheitsfoo.de/census-act/
11. Wikipedia: Global surveillance disclosures (2013–present); http://en.wikipedia.org/wiki/Global_surveillance_disclosure; last visited 2014-08-14
12. Kai Rannenberg: Multilateral Security – A concept and examples for balanced security; Pp. 151-162 in: Proceedings of the 9th ACM New Security Paradigms Workshop 2000, September 19-21, 2000 Cork, Ireland; ACM Press; ISBN 1-58113-260-3

13. Josep Balasch, Alfredo Rial, Carmela Troncoso, Christophe Geuens, Bart Preneel, and Ingrid Verbauwhede: PrETP: Privacy-Preserving Electronic Toll Pricing. Pp. 63-78 in Proceedings of the 19th USENIX Security Symposium, USENIX, 2010
14. Carmela Troncoso, George Danezis, Eleni Kosta, Josep Balasch, and Bart Preneel: PriPAYD: Privacy Friendly Pay-As-You-Drive Insurance. Pages 742-755 in IEEE Transactions on Dependable and Secure Computing - IEEE TDSC 8(5), IEEE, 2011
15. ISO/IEC 24760-1:2011 Information technology – Security techniques – A framework for identity management – Part 1: Terminology and concepts, First edition, 2011-12-15, freely available via http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html
16. Ahmad Sabouri, Ioannis Krontiris, Kai Rannenberg: Attribute-based credentials for Trust (ABC4Trust)"; Pp. 218-219 in Simone Fischer-Hübner, Sokratis K. Katsikas, Gerald Quirchmayr (Eds.): Trust, Privacy and Security in Digital Business - 9th International Conference, TrustBus 2012, Vienna, Austria, September 3-7, 2012; Springer Lecture Notes in Computer Science ISBN 978-3-642-32286-0; see also www.abc4trust.eu
17. Volker Hammer, Karin Schuler: "Leitlinie zur Entwicklung eines Löschkonzepts mit Ableitung von Löschfristen für personenbezogene Daten", Version 1.0.2, Stand 25. Oktober 2013; www.bfdi.bund.de/SharedDocs/Publikationen/Arbeitshilfen/DINLoeschkonzeptLeitlinie.pdf
18. Kai Rannenberg: Where Security Research Should Go in the Next Decade. Pp. 28-32 in Willem Jonker, Milan Petkovic (Eds.): Secure Data Management - 10th VLDB Workshop, SDM 2013, Trento, Italy, August 30, 2013, Post-Proceedings; 2014; Springer Lecture Notes in Computer Science 8425, ISBN 978-3-319-06810-7