Digital Services Based on Vehicle Usage Data: The Underlying Vehicle Data Value Chain

Christian Kaiser[1,3](✉), Andreas Festl[1], Gernot Pucher[2], Michael Fellmann[3], and Alexander Stocker[1]

[1] Virtual Vehicle Research GmbH, Inffeldgasse 21a, Graz 8010, Austria
{christian.kaiser,andreas.festl,alexander.stocker}@v2c2.at
[2] TraffiCon – Traffic Consultants GmbH, Strubergasse 26, Salzburg, Austria
pucher@trafficon.eu
[3] University of Rostock, Albert-Einstein-Straße 22, Rostock, Germany
michael.fellmann@uni-rostock.de

Abstract. The quantify-everything trend has reached the automotive sector while digitalization is a still the major driver of innovation. New digital services based on vehicle usage data are being created for different actors and purposes, e.g. for individual drivers who want to know about their own driving style and behavior or for fleet managers who want to find out about their fleet. As a side effect, a growing number of ICT start-ups from outside Europe have entered the automotive market to work on innovative use cases. Their digital services are based on the availability of vehicle data on a large scale. To better understand and capture this ongoing digital change in the automotive sector, we present an extended version of the Vehicle Data Value Chain (VDVC) originally published in Kaiser et al. (2019a) and use it as a model for better structuring, describing and testing digital services based on vehicle usage data. We classify digital services of two projects by using the VDVC in our paper, an intermodal mobility service and a pothole and driving style detection service. Thus, we evaluate the VDVC and show its general applicability and usefulness in a practical context.

Keywords: Big data · Big Data Value Chain · Vehicle Data Value Chain · Digital services based on vehicle usage data · Connected services · Crowdsourcing of data

1 Introduction and Motivation

Modern mobility is an important driver of our increasingly global economy: raw materials are transported around the globe and processed into products in value-added processes until they finally find their way to the customer via many intermediate stations. Passenger cars and trucks are assembled in a complex supply chain consisting of many small parts and components and finally manufactured in several value-added steps on a production line before they are delivered to customers. This basic business principle was very successful in many domains for a long time, until digitalization added another business aspect, which is becoming an important driver and has even become the decisive

A. Bozzon et al. (Eds.): WEBIST 2019, LNBIP 399, pp. 22–43, 2020.
https://doi.org/10.1007/978-3-030-61750-9_2

criterion in many sectors, including the automotive industry (Accenture 2016). Similar to smartphones, where the focus is no longer on the original innovation, i.e. telephoning, but on digital apps, it is becoming increasingly important for vehicles, too, which digital functionalities they offer - from the Bluetooth connectivity with smartphones to Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V) services or to third-party services that someone can install. And in the context of vehicle use, services based on vehicle usage data have the potential to go beyond the usual application focus of quantified self-applications, namely self-optimization, learning about oneself, social comparison and interaction or gaming, as they can even be extended in a life-saving manner. Driving style analysis, for example, is able to detect driver fatigue and distraction (Lechner et al. 2019), two of the most common causes of accidents. Thus, it is crucial for research to explore how digital services based on vehicle data can improve the practice of driving or enable novel applications for other stakeholders and other markets outside the automotive domain (Stocker et al. 2017).

The basis for digitalization in the automotive domain are the ever-increasing amount of vehicle usage data generated (e.g. modern vehicles are increasingly equipped with radar, lidar and video to support ADAS functionalities) and the ever-increasing capacity of information and communication technologies (ICT) to convert this data into business value for different stakeholder groups. These may include individual stakeholders (e.g. vehicle drivers) as well as organizational stakeholders (e.g. car manufacturers, fleet managers, infrastructure maintainers, or traffic planners). Utilizing “up to one hundred on board control units that constantly communicate with each other” (VDA 2016), modern vehicles are already generating big data using in-vehicle sensors. Certain parts of this data are safety-critical and must therefore not leave the car, while the rest can and will be used for the establishment of novel digital services based on vehicle usage data, which can go far beyond ensuring driving functionality and safety and opens up a multitude of possibilities.

As IT companies enter the automotive market with their services, the balance of power between the players in the automotive industry may also change. IT start-ups have already created several interesting digital services based on data from the vehicle’s on-board diagnostic (OBD) interface or from the driver’s smartphone (Kaiser et al. 2017). This has led to the emergence of new business models in the automotive sector and has even attracted the attention of car manufacturers already. A prominent example is BMW i Ventures and its recent investments in start-ups such as Nauto (improving the safety of commercial fleets, investment made in 2017) and Zūm (providing technologies for reliable child transportation, investment made in 2019). We have now reached the point where it is decided how to go on: Either the large vehicle manufacturers will buy in/redevelop the most promising digital services of the start-ups, or, to the vehicle could merely become an exchangeable device/platform on which digital services run, similar to the smartphone.

Digital services based on vehicle usage data are data processing services which, among other things, work with data related to vehicle driving and can offer added value to users. In this context, the term ‘service’ can be viewed from two different angles: On the one hand, a ‘service’ is understood as a piece of software applying approaches from computer science to transform and merge different sources of data (be it raw data or pre-processed data) into new, enriched forms of aggregated data (Lechner et al. 2019). When performed correctly, the value of these enriched data is inherently higher than the sum of values of the single datasets which were combined in the process. On the other hand, a ‘service’ is understood as something of economic relevance, providing an added value as a service offering to one or more stakeholder groups.

However, the market entry of start-ups has already created a new *data-driven service ecosystem* in the automotive sector, leading to new data flows and collaborations in service development, as Kaiser et al. (2019b) describes. In the high-level view of this empirically obtained ecosystem with experts from the field, there is a data flow from data providers to service providers, who offer services on the market that are consumed by service consumers at the end of the value chain. On closer inspection, for example, there are five ways in which a service provider can already obtain relevant data on a car trip, i) from a market place (e.g. otonomo.io), ii) directly from the car manufacturer (e.g. BMW), iii) from data intermediaries (e.g. HERE Technologies, which has a close relationship to BMW and Daimler), iv) from the results of other service providers and v) from external data sources (e.g. weather services, congestion warnings).

The enormous amount of data available today makes the creation of valid digital services possible in the first place, but also poses a major challenge with regard to data processing (Xu et al. 2017). To create value, data must be acquired, transformed, anonymized, annotated, cleaned, normalized, aggregated, analyzed, appropriately stored and finally presented to the end user in a meaningful way. This implies that an entire data value chain must be created, implemented and monitored. With this in mind, Kaiser et al. (2019a) derived the *Vehicle Data Value Chain (VDVC)* from the Big Data Value Chain as described by Curry (2016) and a literature review on relevant concepts for digital services based on vehicle usage data, including Quantified Self, Big Data, and the Internet of Things. This VDVC is intended to provide a structure and a framework allowing to systematically describe the transformation of data into valuable services, to compare existing digital vehicle services with each other and to understand and explain the data-related challenges associated with them. Hence, the VDVC was used to analyze, summarize and provide insights into existing start-up and vehicle manufacturer initiatives on the market. As a result, we decided to apply the VDVC in the development of services in two case studies, the intermodal mobility service MoveBW (case A) and a pothole and driving style detection service (case B). Finally, this paper is an extended version of Kaiser et al. (2019a), elaborating the VDVC and using another case study of a digital service based on vehicle data for evaluating the improved VDVC.

During the development of digital services based on vehicle data it will always be necessary to obtain an overview of certain characteristics of the individual data value chain steps, e.g. the scope of each step, the input data received in a particular step, the output data generated in a step, typical actors involved, typical architectures that are relevant, relevant trends and tools and, finally, the contribution of a particular step to value creation. For this reason, we subsequently extend the VDVC presented in Kaiser et al. (2019) by adding relevant characteristics to each data value chain step and thus aim to answer the following research question: *What are the relevant steps in developing digital vehicle services that should be part of a data value chain and how can the contribution to value creation be described with characteristics?*

After this introduction and motivation in Sect. 1, our paper continues with a review of background information in Sect. 2. In Sect. 3, we present and describe the extension of the Vehicle Data Value Chain. We then apply this value chain to analyze the intermodal mobility service MoveBW (case A) as well as a pothole and driving style detection service (case B) in Sect. 4. Finally, we draw a conclusion and an outlook of the paper in Sect. 5.

2 Background

2.1 Data as Business Enabler

Tim O'Reilly formulated his extensively quoted principles of Web 2.0 (O'Reilly 2005) including one about the emerging value of data more than a decade ago. Since then, the hype on how to generate added value from all kinds of available data has continued to grow. Data has become the new buzzword. A book by Mayer-Schoenberger and Cukier (2013) on how Big Data is changing our world has become an international bestseller and been cited by researchers more than 5360 times according to Google Scholar. Big Data has received considerable attention from multiple disciplines, including information systems research (Abbasi et al. 2016) and database management (Batini et al. 2015), to name but two.

Due to the exponential growth in the amount of data, for example, an amount of 16 ZB (16 trillion GB) of useful data is expected in 2020 (Turner et al. 2014). It is just a logical consequence that data generation, data analysis, data usage – and the new business models associated with it – have found their way into all areas of life. Homes are increasingly equipped with smart meters, a replacement for mechanic measurement of electricity usage, enabling the emergence of digital services to assist home monitoring and to optimize electricity usage. Smartwatches can track the wearer's movements and, create behavioral data and calculate periodic statistics such as daily, weekly, or monthly walking distances including burned calories per day, week, or month. Many people use their smartphones when exercising to gather extra information on their workout effectivity.

Smartphone apps such as Runtastic (2017a) and Strava (2017) help to monitor how and where people run or cycle, automatically calculating route, pace and periodic statistics including mean speed, time per kilometer, and calories burned. These apps even allow sharing the aggregated data via social networks, thus enabling benchmarking with peers and increasing the joy of exercise. The pattern of collecting, analyzing, and sharing data constitutes the baseline for individual improvements. Instantly calculated and visualized behavioral statistics are easy to compare or share with peers on social media. The collected information per se is not new to these communities. For instance, experienced runners started comparing their real and average time per kilometer using stopwatches a long time ago. However, the simplicity of digital services and the fact that many friends on social media regularly post about their exercising routine has motivated a whole digital generation to track themselves, as 300 million downloads of the Runtastic app (recently renamed to Adidas running) demonstrate (Runtastic 2020). 30 million app sessions per month in Europe produce a reasonable amount of big movement data, which is sufficient for performing representative data analyses and have led to an acquisition by the sports clothing company Adidas. To summarize, digitalization has greatly simplified data collection and analysis methods which used to be too complex and/or only available to experts. Hence, more and more people are joining the self-tracking movement and, in turn, produce more and more data which can be exploited using novel digital services.

2.2 A Value Chain for Big Data

In contrast to all previous technical or organizational innovations, the Internet age has made it possible for data volumes to reach undreamt-of dimensions. Big Data refers to the current conglomerate of newly developed methods and information technologies to capture, store and analyze large and expandable volumes of differently structured data. In a definition by Demchenko et al. (2013), the defining properties of Big Data are Value, Variety, Velocity, Veracity and Volume as shown in Fig. 1. Exploiting the new flows of data can even improve the performance of companies, if the decision-making culture is appropriate (McAfee and Brynolfsson 2012).

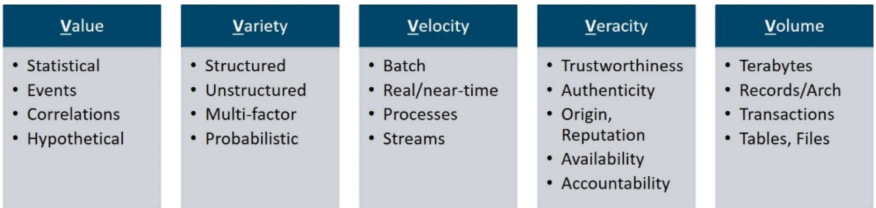

Fig. 1. The 5 vs of big data (Demchenko et al. 2013).

It seems that smart things are increasingly based on big data analysis, which makes it possible to speak of an intimate relationship between those two. While in the Web 2.0 era data was mainly generated by humans sharing user-generated content on portals including YouTube, Wikipedia, or Facebook, the Internet of Things has led to new patterns of data generation driven by machines. Smart, connected objects equipped with all kinds of sensors have now taken over this task (Porter and Heppelmann 2014 and 2015). The Quantified Self phenomenon is making use of these data generated by things (Swan 2009, 2015). Quantified Self refers to the intention to collect any data about the self that can be tracked, including biological, physical, behavioral, and environmental information. Making use of these data to establish applications and services has become a major creator of value. This value is created through multiple activities which are chained together, while the value of the output is steadily increasing.

A company's activities to create and build value were once described by Porter and Millar (1985) with the so-called concept of the value chain. However, this value chain concept can be applied to the data domain to describe activities ranging from data generation to the usage of data in data-driven services for the customer. Data value chains are a model to describe data flows as a series of steps, each of them transforming the value of data. Recently, Åkerman et al. (2018) described a data value chain in the context of production, where data analytics leads to regulations of a production system like in a closed loop control system. Furthermore, the concept of data value chains has been used to describe the value of Linked Data (Latif et al. 2009) and Big Data (Curry et al. 2014) as illustrated in Fig. 2. As modern vehicles are likely to produce big data (e.g. from and for (semi-)automated vehicles), the Big Data Value Chain including several steps of Big Data transformation in the process of generating the data-driven result with the maximum business value is of high relevance to the automotive sector (Xu et al. 2017).

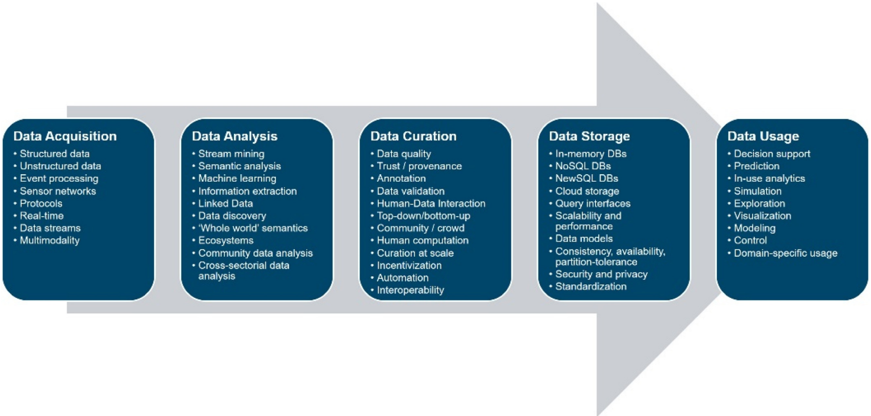

Fig. 2. The big data value chain of Curry et al. (2014)/Curry (2016).

2.3 Big Data Based on Vehicle Usage Data

The automotive industry is also constantly finding innovations for its vehicles as a result of electrification and comfort requirements. For example, mechanical components such as hand brakes or window lifters are increasingly being changed to electronic versions, such as the electric hand brake and electric window lifters. The status (handbrake is applied or released) and its process status (handbrake is applying/releasing) can be captured and used as input for vehicle safety checks and other features. An applied handbrake will automatically be released if the driver starts driving to prevent damage. The data generated through all these vehicle functions can be captured and used within other scenarios, e.g. to create statistics on how often a window is opened/closed or how often somebody is wedged in.

Also due to the common practice of vehicle development to purchase many components from suppliers, many vehicle sensors have so far only been used to provide and support a specific functionality and to increase comfort and safety, although these vehicle sensor data may also be interesting for third parties. As sensors and car features may widely differ from manufacturer to manufacturer and even per car variant, there is not only one single truth about how much data is effectively generated by a modern vehicle today. For instance, the participants from the European research project AutoMat state in a deliverable (Automat 2018) that about 4000 CAN bus signals (one signal could be one measurement value) per second create up to 1 GB of data per CAN bus (without mentioning a sample rate). According to Pillmann et al. (2017), there are “usually 4–12

CAN busses in one car" (with varying amounts of input signals). This clearly shows the high amount of data generated as a by-product during vehicle use.

For highly automated driving, several camera, radar and LiDAR (Light Detection and Ranging) systems are currently being implemented in the vehicles to cover every corner of the vehicle environment. Autonomous vehicles may be forced to exchange information with other vehicles (V2V) and with the infrastructure (V2I), which will boost the amount of available vehicle data enormously in the future. Considering different countries and different patterns of individual driving and mobility behavior, bringing highly automated driving into practice will be a grand digitalization challenge.

Although only part of this data is available for digital vehicle services (e.g. the high sampling rates generate such large amounts of data that the limits of data transmission are exceeded, which would require re-sampling at a lower rate or some signals are simply not relevant) and while only a portion of these data will be made accessible due to safety reasons (EU 2013), the remainder of accessible sensor data from modern vehicles will most likely be sufficient to design and develop a reasonable number of novel digital vehicle services for various stakeholder groups, including individual drivers, various organizational customers, government authorities, and the automotive industry (Kaiser et al. 2017). To sum up, modern vehicles already constitute impressive generators of big vehicle usage data.

3 A Value Chain for Vehicle Usage Data

3.1 Quantified-Self

Digital natives like to have access to services anytime and anywhere and are therefore willing to let their mobile devices such as smartphones and smart watches generate data around the clock. Increasing the knowledge about oneself and eventually enabling new discoveries while performing physical activities including running or cycling has turned into a business-relevant phenomenon. The behavior of turning collected data about oneself into actionable knowledge and insight which is valuable for other stakeholders, too, has been termed Quantified Self. Interestingly, the quantified self-phenomenon has recently been successfully transferred to the automotive industry by US-based start-ups. In this sense and quite analogously, Quantified Vehicles (Stocker et al. 2017) imply a successful transformation of data from different kinds of sensors related to the vehicle (in-vehicle sensors, smartphone and wearable sensors used by the driver) into actionable knowledge, e.g. on the behavior of the vehicle. This way, they generate value for different kinds of stakeholders that are part of digital vehicle data service ecosystems such as insurance or fleet management providers, finally resulting in novel digital services based on vehicle data in various domains (Kaiser et al. 2018b, 2019b).

Self-tracking with consumer devices, as shown in the example of Runtastic (Adidas running), can also be transferred to vehicles: Vehicles already collect a large amount of operating data via sensors and control units that ensure the functionality of the vehicle. However, these big vehicle data could be used to enable a series of apps and services for different customer groups. The market value for vehicle usage data is considered to be even higher than for other markets due to the importance of vehicles in first world countries. A number of US-based ICT start-ups seized this opportunity, now offering smartphone and web applications providing insights into vehicle-generated data, after they received up to €25 million of funding from investors (Stocker et al. 2017). Interestingly, while some car manufacturers and suppliers (e.g. Magna International, Continental ITS, and BMW i Ventures) are among the investors, forming strategic partnerships with start-ups, others participate in research projects and try to keep data-related business in their own area of influence. This holds for Volkswagen, for example, which coordinates the EU project AutoMat to develop a marketplace for vehicle lifecycle data (Stocker and Kaiser 2016). Furthermore, recent reports from the German automotive industry association (VDA) suggest that car manufacturers "have to hold a stronger position in the future and may limit the capabilities of third parties to freely access car data." To summarize, the potential of vehicle usage data seems to be such that it has become a battle worth fighting (Kaiser et al. 2017). How vehicle usage data generates value leads us to the next section in which we describe the Vehicle Data Value Chain.

3.2 The Vehicle Data Value Chain (VDVC)

To systematically describe the transformation of data into valuable services, the concept of value chain can create a suitable structure and framework. In this regard, we propose the Vehicle Data Value Chain (VDVC) as a lightweight model. We derived the VDVC from the Big Data Value Chain (Curry 2016, illustrated in Fig. 2). We adapted Curry's value chain regarding the name, number and order of stages to reflect our experiences from research projects in the automotive domain. The stage of *Generation* (of vehicle usage data) was added as a separate stage to explicitly reflect the origin of the data (e.g. in-vehicle or related sensors). The stage *Acquisition* (of vehicle usage data) corresponds to Curry's Data Acquisition. Moreover, we have changed the order of Curry's stages of analysis and curation since we interpret the terminology differently. For example, Curry seems to include normalization procedures implicitly within machine learning in the stage of Data Analysis, whereas we consider this as an important separate pre-processing step which correlates with Curry's stage of Data Curation. Hence, we have re-named Curry's stage of Data Curation, *Pre-processing*, which is followed by the stages *Analysis*, *Storage*, and *Usage* (in each case: of vehicle usage data), as visualized in Fig. 3. As the result of the processing could be the input for further analysis, an arrow back to Acquisition indicates the possible a circular path.

Furthermore, to compare digital services based on vehicle data and to understand and explain the data-related challenges associated with them, we added eight characteristics to each value chain step: *i) Description/Scope* to describe the scope of the step, *ii) Input examples* and *iii) Output examples* to name possible inputs and outputs per step, *iv) Actor examples* to name relevant actors in this step, *v) Architecture examples* to describe which architecture usually is used in a specific step, *vi) Trend examples* to name current trends in the specific value chain step, *vii) Tool examples* to name possible tools and *viii) Contribution to value creation* to summarize the contribution of this step to value creation. The single value chain steps are shown in Fig. 4 and are described in the following subsections.

Generation (of Vehicle Usage Data). This step has the scope of generating measurements through any sensors which can capture condition data directly (engine RPM or

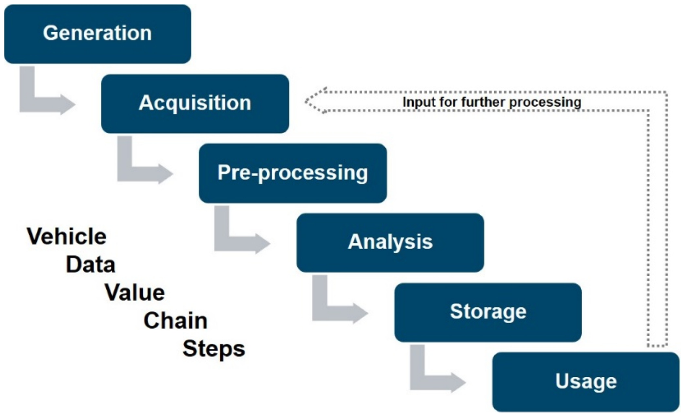

Fig. 3. The vehicle data value chain derived from Curry (2016) and based on Kaiser et al. (2018b) and Kaiser et al. (2019a).

vehicle speed) or indirectly (road surface). In the case of direct influence, we see three main data sources: In-vehicle sensors, smartphone sensors and sensors in individual user devices (e.g. a pulse watch). An indirect data sources can be literally any data source that provides information on the state of a vehicle, its driver or surroundings; an example could be a road operator camera to display traffic flow. This process step is essential for the vehicle data value chain, since the data origin determines the reliability and the type of influence (direct, indirect). The current trend to equip modern vehicles with ADAS functionalities (e.g. through the use of radar and lidar sensors for better detection of the driving environment) increases the amount of data generated and the possibilities for use cases once more.

Acquisition (of Vehicle Usage Data). This step describes the process of collecting the generated data. In-vehicle sensor data is not directly accessible as it is secured in order to safeguard vehicle functionality and is therefore only exchanged between the various electronic control units via one of the vehicle's internal bus systems, e.g. CAN bus. However, a filtered quantity of this sensor data is accessible via the On-board diagnostic (OBD) interface (Turker and Kutlu 2015), which is intended to be used by service staff to read out the generated error messages. It is therefore possible to develop plug-in devices with an internet connection, thereby effectively using the OBD-port as a source of sensor data. There are already some professional solutions with data acquisition devices built into the vehicle, which read signals directly from the CAN bus in an unfiltered way. To meet the requirements of the EU Directive 2010/40/EU – establishing inter alia the costless provision of universal, road safety-related minimum traffic information (EU 2013) - a standardized interface would be feasible sooner or later. Data from smartphone sensors is acquired using specific applications, capable of gathering and transmitting data. In the case of external data sources, the main issues are the varying availability and quality levels of the data. For example, APIs usually limit the number of requests allowed per time interval, so the acquisition process must be adapted to meet these thresholds.

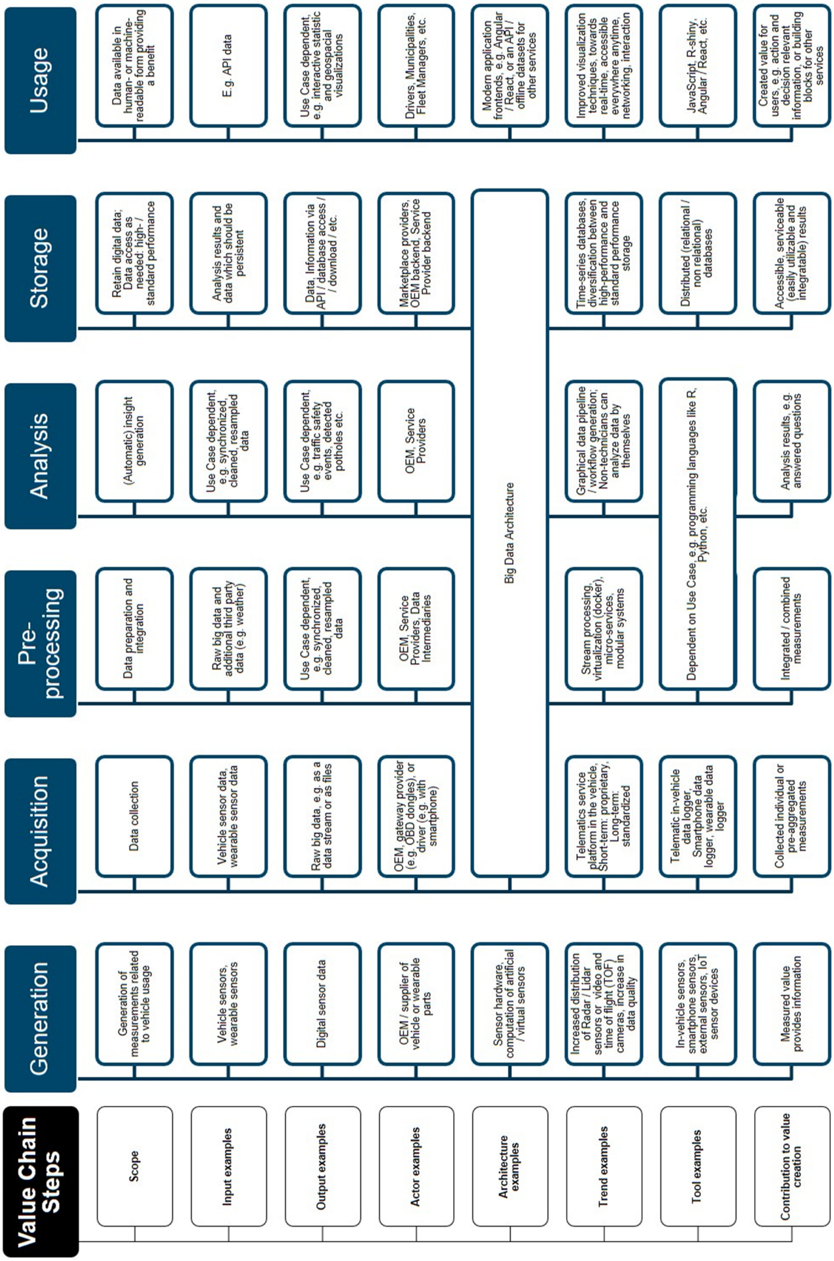

Fig. 4. The vehicle data value chain derived from Curry (2016) and based on Kaiser et al. (2018b) and Kaiser et al. (2019a) extended with characteristics.

Gathered data is stored for further processing; the chosen storage and format heavily depend on the subsequent processing steps.

Pre-processing (of Vehicle Usage Data). This step consists of the process of data preparation and integration. It is the sum of any anonymization, annotation, cleansing and normalization activities before any data analysis is conducted. Sensor values including private user information, erroneous sensor readings, different sensor sampling frequencies or unsynchronized data are examples of issues addressed in this stage. Data quality has a high impact on service quality. For instance, if the accuracy of the GNSS signal is low, a trip may not be linked to the correct road and may lead to wrong conclusions.

Analysis (of Vehicle Usage Data). This step is the process of automatic insight generation, with the purpose of extracting useful hidden information. This involves linking data from different data sources, exploring the data, performing statistical analyses and using machine learning algorithms to detect latent information hidden in the data. For instance, weather data can be linked to vehicle speed on a particular road to determine whether the driver is driving differently in wet or icy conditions. Weather data can be linked to acceleration data to determine whether a driver is driving aggressively in bad weather conditions.

Storage (of Vehicle Usage Data). In this step of the value chain, proper data access is established. It is already defined in the Big Data Value Chain as "the persistence and management of data in a scalable way that satisfies the needs of applications that require fast access to the data" (Curry 2016). In the case of vehicle sensor data, persistent storage is usually achieved by using a combination of classical relational databases (for metadata), Big Data file systems (for raw input data) and so called "time series databases" to store data that changes with time, which allow fast analyses on the stored contents.

Usage (of Vehicle Usage Data). The final step deals with making the data available in human- or machine-readable form (or both, as required). It includes all kinds of user or software interaction with the collected data and any conclusions derived from it in the above-mentioned process. The retrieved data could either be regarded as the end result of the process, being presented more or less directly to end users, or it could serve as input for further processing steps, thus forming a circular path in the processing chain.

4 Evaluation of the VDVC

4.1 Case A: Description of the Intermodal Mobility Service MoveBW

A regional, intermodal mobility service called MoveBW helps to increase the compliance rate of transport users (e.g. the percentage of people using a park and ride option) in relation to the current transport strategy of the region. The strategy offered by an European industry consortium mainly aims at meeting air quality targets and reducing traffic jams all over the federal province of Baden-Württemberg (Germany), including its provincial capital Stuttgart.

Geographically situated in a valley basin, Stuttgart, like all cities in valley basins (e.g. Graz), struggles with air pollution through fine dust. Thus, the city of Stuttgart continuously develops transport strategies to better comply with air quality regulations. In the past, these strategies were communicated to the public using radio traffic messages or electric traffic signs only. However, the compliance rate and thus success were comparably low. The MoveBW mobility service smartphone application aims to increase said compliance rate, especially that of visitors new to the region. It does so by including easy-to-use routing functionalities which are connected to rewards: Bonus points are granted if a user follows the recommended route. Collected bonus points can later be exchanged for immaterial or monetary values.

The intermodal journey planner allows users of the MoveBW smartphone application to plan their trips in advance. They can pick their preferred combination of transport modes from different options suggested to them. Additional information is displayed, not only showing travel time, but also eco-friendliness, travel costs and incentives gained (e.g. public transport vouchers and CO2 savings). Moreover, it is possible to directly book tickets for the different modes of transport included in their preferred journey and yet to receive only one bill. In this way, transport services such as public transportation, car sharing, bike sharing, and parking space management are integrated conveniently, encouraging users to make efficient use of all modes of transport. The application also provides on-trip navigation and information on traffic obstructions such as construction works or accidents.

The MoveBW services are currently being monitored and evaluated in an extensive test phase. Based on the findings, both the digital service and traffic control strategies will be revised, aiming to maximize favored effects on the individual mobility behaviors of traffic participants, for example by applying different strategies for daily commuters and visitors. The smartphone application is planned to be released in the first quarter of 2019. Mock-ups of the current design are shown in Fig. 5.

Fig. 5. The MoveBW smartphone application provides functions for intermodal journey planning, traffic information, ticketing and on-trip navigation. (Source: https://www.altoros.com/blog/mobile-devices-are-propelling-industrial-iot-scenarios/).

A special challenge regarding data management is the multitude of data sources for the intermodal routing algorithms in the MoveBW App. The Vehicle Data Value Chain introduced in Sect. 3 helps to provide a clearer view. Its application to the underlying data transformation process, from Data Generation to Data Usage, is shown in Table 1.

Table 1. An overview of the MoveBW-Service. (Source: Kaiser et al. 2019a).

VDVC step	Description of MoveBW-service
Data generation	Various sensor data and basic reference data is considered, e.g. – floating car data: average mean travel time per road segment based on anonymized GNSS data of vehicles, – stationary traffic measurement: rate of flow for single measurement locations, – public transport: schedule and sometimes occupancy rate, – car park interfaces: occupancy rate, – park & ride interfaces: occupancy rate, – air quality measurement units: air quality measurements and forecast (includes weather forecast);
Data acquisition	Querying web APIs from the various data sources. Additionally, the smartphone App which is described in Data Usage provides GNSS information, which is used for on-trip routing and to detect which means of transport the user actually uses to be able to reward them if the recommended option is used
Data pre-processing	Annotation, normalization and semantic extraction of data. Transformation of data to meet a common reference basis (in this case a public transport grid, no typical geo-coordinates). Furthermore, GNSS data from the smartphone App is anonymized (start- and end-trajectories are truncated). In this step the data is hosted in a distributed database system (e.g. PostgreSQL cluster)
Data analysis	A dynamic routing algorithm which also takes the provided intermodal transport strategy, CO2 savings, and personal preferences into account. A self-developed algorithm which utilizes pgRouting (an open source project to extend PostGIS/PostgreSQL to provide geospatial routing functionality) and the popular Dijkstra algorithm (to find the shortest path between nodes). Provision of routing recommendations (weightings for routes) through this algorithm
Data storage	A distributed database system, e.g. a PostgreSQL cluster
Data usage	The MoveBW App currently being developed should help the commuter to choose a mode of transport and guides the commuter to the selected destination in compliance with environmentally-oriented traffic management strategies

In case of MoveBW, where all steps of the MoveBW service are known to the authors, the VDVC provides a framework to describe the service layer by layer and thus also helps others to understand the service and its underlying value chain.

In the next section, the development of a pothole and driving style detection service is described using the VDVC.

4.2 Case B: Description of a Pothole and Driving Style Detection Service

Generating value out of vehicle data is a challenging task: For this purpose, vehicle data analytics has become an important technique in identifying the value of generated vehicle data. However, to exploit this value in products and services, several steps must be performed, and several (not only technical) challenges have to be solved. In the beginning, an appropriate analytics question must be identified such as e.g. identify the driving style of the driver from vehicle data, detect the road surface quality, identify potholes on roads, or predict the engine's wear.

Then, vehicle data must be captured: Three different approaches for data capturing are possible: the installation/use of own sensors within the vehicle to record vehicle movements and other contextual information, the connection of a vehicle data logger to the vehicle's on board diagnostic (OBD) interface to capture vehicle data such as vehicle speed or RPM, or the installation of a professional Controller Area Network (CAN) logger to obtain even more vehicle data from the vehicle's CPUs such as for example the state of vehicle assistance systems or the steering wheel angle. While the first option is probably the simplest one, it can only record contextual data and track the movement of the vehicle, but it does not allow access to vehicle sensors. The second option can provide already access to some vehicle sensor data such as vehicle speed or engine temperature, which is relevant for testing whether the vehicle's emissions are still within tolerance. The third option in theory provides access to all vehicle sensor signals, but only if the device listening to the CAN bus can decode the streamed raw CAN bus data to readable data, requiring either the vehicle manufacturer or the respective vehicle CPU manufacturer to provide the necessary decoding information (usually referred to as CAN-DBC files) (Fig. 6).

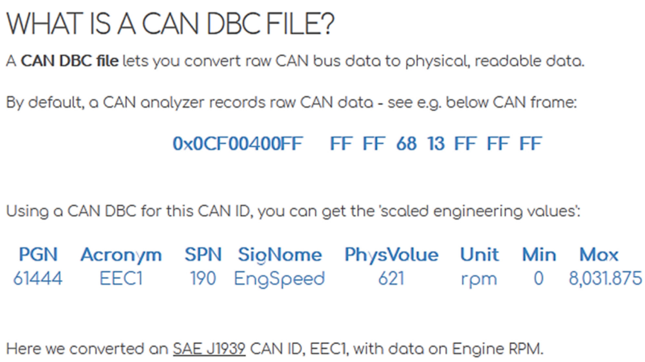

Fig. 6. CAN DBC files. Source: CSS electronics (2020).

Different data loggers may store the data in different formats. Typically, they can collect multiple signals at once, which are all transmitted on the same wire. Thus, the logger needs to know and save at least three different properties of the data: What was measured, what was its value and when was it measured. This naturally leads to a tabular format very similar to the example depicted in Table 2.

Table 2. Vehicle raw data structure (example).

Timestamp	Signal name	Signal value
2019-9-13 5:28:36.206089	RPM	1500
2019-9-13 5:28:36.226331	Acceleration-X	0.476
2019-9-13 5:28:36.245312	Vehicle Speed	39
2019-9-13 5:28:36.268915	Engine oil temperature	90
..	..	..

While this format is convenient for the logger to store data, it is much less suited for a statistical analyses or automated processing of the data. There are three main difficulties: First, several signals are mixed together in one column, creating the need for grouping and filtering even before very simple operations. Second, there can be multiple signals that were measured at the same time, requiring the analyst to investigate multiple rows at once to check a single instance in time. The third difficulty lies in the varying sampling rates of the signals. Each signal may have been captured with a different rate and even within a single signal, smaller deviations of the sampling rate are possible and common. Clearly pre-processing of the captured vehicle data is needed to make it better explorable for data analysts.

After the required vehicle data is stored, a series of further steps must be performed to prepare the data for analysis. This data (pre-)processing process can be quite comprehensive and depends very much on the respective analysis question to be solved, e.g. the detection of potholes from vehicle data. A crucial step in this process is the alignment of the coordinate system of data logger and vehicle. Many signals are vector valued, with acceleration as the maybe most prominent example. To simplify analyses and interpretations, it is highly desirable to express these vectors in the reference frame of the car, i.e. x-Acceleration should be the component in the x-direction of the car/the driving direction. In general, one cannot assume that the logger was mounted such that its internal coordinate system corresponds to the one of the vehicle. This is especially true when cheap devices that are mounted by end-users are used. Any misalignment of the reference frames needs to be detected and corrected prior to analysis.

As with most other data types, vehicle data signals should be searched for missing values, wrong values, and outliers and these should be removed. Some signals may contain a lot of noise and must be smoothed. To separate the signals into different columns the data should be transformed using the 'signal name' as pivot. Simultaneously, it makes sense to resample each signal to a common sampling rate from the analysis' viewpoint. The "right" sampling rate again depends on the question the be answered. The result is than in a similar form as depicted in Table 3. Now each row corresponds exactly to a point in time and the time interval between the rows is constant, in this example 0.1 s/10 Hz.

Table 3. Structure of pre-processed vehicle usage data (example).

Timestamp	Engine speed	Acceleration-X	Vehicle speed	..
2019-9-13 5:28:36.20000	1500	0.477	39	
2019-9-13 5:28:36.30000	1501	0.479	40	
2019-9-13 5:28:36.40000	..	..	..	

The data prepared in this way can now be used to work on the vehicle data analysis question and/or to search for interesting events (such as potholes for example). Depending on the type of event, multiple signals can be relevant. Events should usually be post-processed to combine separate events, which are only divided by a short-time interruption, into a single event. The recorded events may be linked with weather and position data, so that for each event the time and place of occurrence as well as the prevailing weather is known.

For different types of events, different detection methods need to be employed. One can detect a pothole event (driving over a pothole) by investigating acceleration values and rotation rates as follows: Consider the acceleration normal to the road, as well as the vehicle's rotation around its lateral axis ('pitch') The acceleration readings will exhibit a distinct spike, while a certain pattern is simultaneously visible in the rotation rate: When the front tires are in the pothole, the front of the vehicle is lower than the rear, if the rear tires are in the pothole, it is vice versa, causing a rotation around the lateral axis. This results in a typical "pitch" movement that can be detected. In a last step, the results of the analysis - in this case the detected potholes - can be visualized on a map. In our case this supports drivers in not choosing bad roads, or support road operators in better maintaining roads.

To detect strong acceleration and braking events, the signals vehicle speed, acceleration in the direction of travel and rotation around the lateral axis ("pitching") are particularly suitable. The "pitching" is caused by the change in weight distribution when the speed changes: when a vehicle is accelerating, more weight moves to the rear axle - the rear drops and the front rises. When a vehicle is braking, the opposite is true. These movements can be detected. However, since detection using only a single signal can be prone to error, we always use several signals in our algorithm, which must all deflect simultaneously to trigger detection.

The driving styles of drivers can differ in many facets (e.g.: comfort level, gear choice, aggressiveness). Depending on the type of vehicle the driving style may have a large influence on fuel/power consumption, component-wear and road safety. In an attempt to quantify this influence, we use all calculated events to calculate a 'risk score' that indicates how unsafe a single trip was. The more safety-related events per time unit occurred in a trip, the higher the value. Furthermore, we consider the influence of environmental conditions in our calculations. For example, heavy braking in rain will result in a higher risk than the same braking on a dry road. To make the risk score interpretable, we normalize it using the scores from all available trips as a basis. We then present the value as statistical rank, for example a value of 56.72% means that this trip is safer than 56.72% of all trips in the database. In a map visualization, the driver is presented the trip with markers indicating start and stop positions, as well as locations for safety-relevant events.

Based on this methodology, a smartphone application, shown in Fig. 7, has been developed for drivers interested in monitoring their driving style.

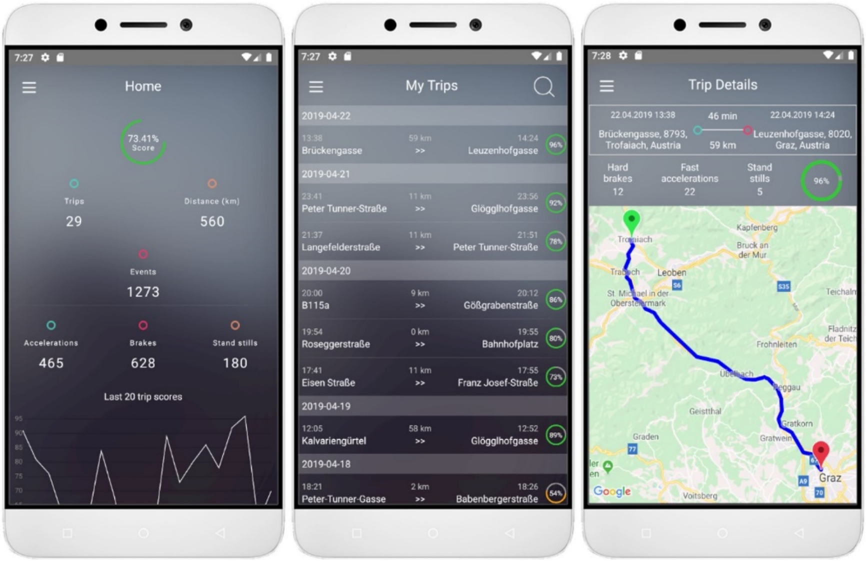

Fig. 7. A smartphone application for driving style detection.

On the left screen named *Home*, the driver has an overview of his trips. In the presented figure, his overall score is 73.41%; he has 29 trips with a total distance of 560 km. In these trips 1273 events have been detected, which are composed of 465 acceleration events, 628 brake events and 180 stand-still events. On a second screen named *My Trips,* which is displayed in the center, a list of the most recent trips, grouped by date, is shown. For each trip, the information on which location and at which time the trip started and ended is displayed together with the trip score and the trip distance. Selecting one of the trips opens a third screen named *Trip Details*, where additionally the events are decomposed into categories and the trip is visualized on a map.

Now that we have described the idea of this service, we want to show in the following table how clear and comparable the service becomes by using the VDVC (Table 4).

Table 4. An overview of the pothole and driving style detection service.

VDVC step	Description of pothole and driving style detection service
Data generation	Vehicles are equipped with data loggers that record the signals required for pothole and driving style detection (e.g. speed, acceleration, rotation, position, etc.). These data loggers are connected to the on-board diagnostic interface of the vehicle and additionally generate acceleration, rotation and GPS data
Data acquisition	Vehicle movement data including OBD measurements as well as acceleration, rotation and position measurements is periodically recorded and imported as raw vehicle data into a local PostgreSQL database on the data logger. The collected data is made available as a data stream or as manually exported files in a PostgreSQL database running in the cloud
Data pre-processing	The pre-processing of the vehicle data includes the alignment of the datalogger's coordinate axis with the trajectories of the vehicle, the search for missing and incorrect values and outliers and their elimination, the smoothing of the signals to reduce noise and the interpolation of all signals to a useful sampling rate. Additionally, contextual weather data is integrated
Data analysis	For pothole detection, the acceleration perpendicular to the road and the "pitching" of the vehicle (i.e. the rotation around the transverse axis) are used. If these exceed certain threshold values, a pothole event is generated. In comparison, vehicle speed, acceleration in the direction of travel and rotation around the transverse axis ("nodding") are used to detect events relevant to driving safety, such as strong acceleration, braking and cornering maneuvers. If these exceed certain threshold values, a harsh acceleration, braking and cornering event occurs
Data storage	The events calculated in the analysis phase (harsh acceleration, braking, cornering as well as potholes) are stored in the PostgreSQL database together with their GPS locations and the corresponding weather information to visualize them on maps and perform additional statistical analyses, such as calculating a risk score for a single trip, taking into account the amount and severity of detected events per trip length as well as the respective weather conditions and a cumulative risk score for a driver
Data usage	Drivers should be provided with information to improve their driving. The application shown in the figure above should help the driver to monitor his own driving and compare it with the driving of other drivers in order to improve driving safety. Finally, the application can visualize detected potholes so that the driver can avoid driving into these potholes

5 Conclusion and Outlook

An increasing number of digital services based on vehicle usage data are offered on the market and are increasingly used and demanded by users. Digitalization has not only become an important driver of innovation in the automotive sector, but may also change the balance of power in the automotive sector in the long term. With the background that our society is strongly driven by mobility, it is almost our duty to examine the emergence of digital services based on vehicle usage data more closely. Consequently, in our paper we have looked at a way of better describing and structuring digital services based on vehicle usage data. After a comprehensive analysis of related work, we have reviewed two different digital services by using the VDVC for a better structured description of how value is created. Using the VDVC model, we explicitly describe which activities must be carried out in the individual steps of the value chain in order to finally enable these two services.

As an outlook, it should be mentioned that digital vehicle services and the required technological infrastructure to facilitate data acquisition, pre-processing, analysis and storage, are currently a hot topic in the automotive domain. There are already ideas for using blockchain technology and brokers to make data sharing more transparent and secure, as described in Kaiser et al. (2019). Yet, while some car manufacturers invest in start-ups, others limit access to data via the OBD interface, arguing that they are not suitable for digital vehicle services (VDA 2016; ACEA 2016). In contrast, the European Automobile Manufacturers Association ACEA promotes car data sharing (ACEA 2017).

Regulation (EU) No. 886/2013 (concerning the Directive 2010/40/EU on Intelligent Transport Systems ITS), published by the European Commission, has actually been regulating the provision of universal, road-safety relevant minimum traffic information to users free of charge for years and calls on car manufacturers to make safety-relevant data available to the public via national contact points (EU 2013). While the vehicle manufacturers have long referred to the no longer up-to-date transmission standard based on WLAN technology (e.g. G5), several EU-wide initiatives (such as the C-ROADS initiative) have not given up, extending the development to telecommunications technologies (e.g. 4G, 5G) and presenting a concrete implementation plan for C-ITS services with Day 1 Applications. Since the end of 2019, the latest Volkswagen Golf is the first series-production vehicle on the market to use this data exchange standard. The C-ROADS initiative of several EU member states and road operators aims to use C-ITS services to enable the transmission of infrastructure information (e.g. roadside units) to the vehicle cockpit, e.g. to inform about dangerous situations, e.g. a vehicle backing out or pedestrians in the crosswalk behind the next bend. (C-ROADS 2017).

At the same time the International Organisation for Standardisation (ISO 2017) has set up a standardization work group titled *ISO/TC 22/SC 31/WG 6 Extended Vehicle/Remote diagnostics* to inter alia define access, content, control and security mechanisms for the provision of vehicle data for web services (VDA 2016).

Additionally, current EU-funded projects such as EVOLVE are developing solutions to ease the integration and fusion of multiple data sources for the purpose of service and business development using Linked Data (EVOLVE 2019; Latif et al. 2009). “Linked data is a lightweight practice for exposing and connecting pieces of data, information, or knowledge using basic web standards. It promises to open up siloed data ownership and is already an enabler of open data and data sharing” (Rusitschka and Curry 2016).

To conclude, we expect the market of digital services based on vehicle usage data to grow tremendously in the future, as the combination of vehicle data with data from external sources (e.g. weather data, traffic data, open data) will enable new scenarios for digital vehicle services.

Acknowledgement The EVOLVE project (www.evolve-h2020.eu) has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825061. The document reflects only the author's views and the Commission is not responsible for any use that may be made of information contained therein .

References

Abbasi, A., Sarker, S., Chiang, R.H.: Big data research in information systems: toward an inclusive research agenda. J. Assoc. Inf. Syst. **17**(2) (2016). http://ahmedabbasi.com/wp-content/uploads/J/AbbasiSarkerChiang_BigData_JAIS_2016.pdf

Accenture, Digital Transformation of Industries: Automotive Industry (2016). https://www.accenture.com/t20170116T084448__w__/us-en/_acnmedia/Accenture/Conversion-Assets/WEF/PDF/Accenture-Automotive-Industry.pdf. Accessed 08 Jan 2020

ACEA (European Automobile Manufacturers Association): ACEA Position Paper: Access to vehicle data for third-party services (2016)

ACEA (European Automobile Manufacturers Association) (2017). http://cardatafacts.eu/. Accessed 08 Jan 2020

AutoMat. http://automat-project.eu/. Accessed 08 Jan 2020

AutoMat. AutoMat Deliverable D5.3: Full Prototype of Cross-Sectorial Vehicle Data Services (2018)

Åkerman, M., et al.: Challenges building a data value chain to enable data-driven decisions: a predictive maintenance case in 5G-enabled manufacturing. Procedia Manuf. **17**, 411–418 (2018)

Batini, C., Rula, A., Scannapieco, M., Viscusi, G.: From data quality to big data quality. J. Database Manag. **26**(1), 60–82 (2015)

Curry, E., Ngonga, A., Domingue, J., Freitas, A., Strohbach, M., Becker, T.: D2.2.2. Final version of the technical white paper. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4) (2014)

CSS electronics (2020). https://www.csselectronics.com/screen/page/dbc-database-can-bus-conversion-wireshark-j1939-example/language/en. Accessed 16 Jan 2020

Curry, E.: The big data value chain: definitions, concepts, and theoretical approaches. In: Cavanillas, J.M., Curry, E., Wahlster, W. (eds.) New Horizons for a Data-Driven Economy, pp. 29–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-21569-3_3

C-ROADS. Detailed pilot overview report (2017). https://www.c-roads.eu/fileadmin/user_upload/media/Dokumente/Detailed_pilot_overview_report_v1.0.pdf. Accessed 08 Jan 2020

Demchenko, Y., Grosso, P., de Laat, C., Membrey P.: Addressing big data issues in scientific data infrastructure. In: 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, pp. 48–55 (2013). https://doi.org/10.1109/cts.2013.6567203

EVOLVE (2019). https://www.evolve-h2020.eu/. Accessed 08 Jan 2020

EU (2013). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32013R0886. Accessed 08 Jan 2020

ISO (2017). https://www.iso.org/committee/5383568.html. Accessed 08 Jan 2020

Kaiser, C., Festl, A., Pucher, G., Fellmann, M., Stocker, A.: The vehicle data value chain as a lightweight model to describe digital vehicle services. In: 15th International Conference on Web Information Systems and Technologies (2019a)

Kaiser, C., Stocker, A., Viscusi, G., Festl, A., Moertl, P., Glitzner, M.: Quantified cars: an exploration of the position of ICT start-ups vs. car manufacturers towards digital car services and sustainable business models. In: Proceedings of 2nd International Conference on New Business Models, pp. 336–350 (2017)

Kaiser, C., Steger, M., Dorri, A., Festl, A., Stocker, A., Fellmann, M., Kanhere, S.: Towards a privacy-preserving way of vehicle data sharing – a case for blockchain technology? In: Dubbert, J., Müller, B., Meyer, G. (eds.) AMAA 2018. LNM, pp. 111–122. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-99762-9_10

Kaiser, C., Stocker, A., Festl, A., Lechner, G., Fellmann, M.: A research agenda for vehicle information systems. In: Proceedings of European Conference on Information Systems (ECIS) 2018 (2018b)

Kaiser, C., Stocker, A., Fellmann, M.: Understanding data-driven service ecosystems in the automotive domain. In: Proceedings of Americas Conference on Information Systems (AMCIS 2019) (2019b)

Latif, A., Saeed, A.U., Hoefler, P., Stocker, A., Wagner, C.: The linked data value chain: a lightweight model for business engineers. In: Proceedings of I-Semantics 2009. 5th International Conference on Semantic Systems, pp. 568–577 (2009). Journal of Universal Computer Science

Lechner, G., Fellmann, M., Festl, A., Kaiser, C., Kalayci, T.E., Spitzer, M., Stocker, A.: A lightweight framework for multi-device integration and multi-sensor fusion to explore driver distraction. In: Giorgini, P., Weber, B. (eds.) CAiSE 2019. LNCS, vol. 11483, pp. 80–95. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21290-2_6

Mayer-Schoenberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, Boston (2013). ISBN 0544002695 9780544002692

McAfee, A., Brynjolfsson, E.: Big data: the management revolution. Harv. Bus. Rev. **90**, 60–68 (2012)

Nauto. https://www.nauto.com/. Accessed 08 Jan 2020

O'Reilly, T.: What is web 2.0. O'Reilly Media, Sebastopol (2005)

Pillmann, J., Sliwa, B., Schmutzler, J., Ide, C., Wietfeld, C.: Car-to-cloud communication traffic analysis based on the common vehicle information model. In: IEEE Vehicular Technology Conference (VTC-Spring) Workshop on Wireless Access Technologies and Architectures for Internet of Things (IoT) Applications (2017)

Porter, M.E., Millar, V.E.: How information gives you competitive advantage (1985)

Porter, M., Heppelmann, J.E.: How smart, connected products are transforming competition. Harv. Bus. Rev. **92**, 64–88 (2014)

Porter M., Heppelmann J.E.: How smart, connected products are transforming companies. Harv. Bus. Rev. **93**, 96–114 (2015)

Runtastic (2017a). https://www.runtastic.com/en. Accessed 08 Jan 2020

Runtastic (2020). https://www.runtastic.com/career/facts-about-runtastic/. Accessed 19 Jan 2020

Rusitschka, S., Curry, E.: Big data in the energy and transport sectors. In: Cavanillas, J.M., Curry, E., Wahlster, W. (eds.) New Horizons for a Data-Driven Economy, pp. 225–244. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-21569-3_13

Stocker, A., Kaiser, C.: Quantified car: potentials, business models and digital ecosystems. E & i Elektrotechnik und Informationstechnik **133**(7), 334–340 (2016)

Stocker, A., Kaiser, C., Fellmann, M.: Quantified vehicles. Bus. Inf. Syst. Eng. **59**(2), 125–130 (2017)

Strava (2017). https://www.strava.com. Accessed 08 Jan 2020

Swan, M.: Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int. J. Environ. Res. Public Health **6**(2), 492–525 (2009). https://doi.org/10.3390/ijerph6020492

Swan, M.: Connected car: quantified self becomes quantified car. J. Sens. Actuator Netw. **4**(1), 2–29 (2015)

Turker, G.F., Kutlu, A.: Methods of monitoring vehicle's CAN data with mobile devices. Glob. J. Comput. Sci. **5**(1), 36–42 (2015). http://dx.doi.org/10.18844/gjcs.v5i1.31

Turner, V., Gantz, J.F., Reinsel, D., Minton, S.: The digital universe of opportunities: rich data and the increasing value of the internet of things. Rep. from IDC EMC (2014)

VDA: Access to the vehicle (and vehicle generated data) (2016). https://www.vda.de/en/topics/innovation-and-technology/network/access-to-the-vehicle.html. Accessed 08 Jan 2020

Xu, W., Zhou, H., Cheng, N., Lyu, F., Shi, W., Chen, J., Shen, X.: Internet of vehicles in big data era. IEEE/CAA J. Autom. Sin. **5**(1), 19–35 (2017)