Business Models for Smart Mobility IoT Data

Reinier Klaas de Ridder
University of Twente
Drienerlolaan 5, 7522 NB Enschede
The Netherlands
r.k.deridder@student.utwente.nl

ABSTRACT

The Internet of Things (IoT) has shifted from a conceptional and developmental phase to an implementation and exploitation phase. One of the sectors where this is the case is road transportation. Connected vehicles are said to offer major contributions in solving the problems of sustainability and overutilization of infrastructure. A lack of well-defined value propositions halts the adoption of this potent technology. This paper identifies factors that are important in the data monetization decision making process for vehicle manufacturers. These factors are derived from an analysis of the collected data and possible stakeholders. This paper stresses the importance of the use of public data marketplaces to accommodate most possible use cases and data business models.

Keywords

Internet of Things, Transportation, Business Model, Data Exploitation, Smart Mobility

1. INTRODUCTION

The Internet of Things (IoT) is part of the latest technological advancement of this century. Alongside Blockchain and Big Data, it became one of the most 'hyped' concepts in business environments. Now, real applications and benefits are displacing the hype. The question 'What is IoT?' is transformed to 'How should IoT be used?'. According to Gartner, this *"shift is largely fuelled by the disruptions in IoT technology that have occurred, enabling a feasible and achievable path to creating business value"* [14].

Mark Wieser (1991) was one of the first to lay out a contemporary vision of the IoT. In his paper on ubiquitous computing, he starts with the words: *"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it"* [31]. Till date, there is no common agreement on the formal definition of the IoT. Several proposals have been made, however. Haller et al., for example, proposed: *"A world where physical objects are seamlessly integrated into the information network, and where the physical objects can become active participants in business process"* [16].

31th Twente Student Conference on IT July 5th, 2019, Enschede, The Netherlands.

1.1 Challenges for IoT

While the technology now seems finalised for widespread adoption, several challenges are yet to be faced. One of these challenges is so called platform fragmentation and a severe lack of technical standards. This means that IoT devices are developed to operate within a specific ecosystem, with little to no horizontal compatibility. As a consequence, businesses are forced to adopt this given ecosystem to prevent incompatibility issues [11]. This approach also slows the ability to create applications that operate between these inconsistent ecosystems.

Secondly, several privacy and security concerns have risen. IoT devices are capable of continuous collection of data, personal and non-personal. When widely adopted, this collection will often happen without consent, possibly leading to major invasions of privacy. Not only due to the lack of consent, but also due to the absence of thorough data-ownership policies [19]. This danger is enlarged by numerous security issues in the current IoT space. These vulnerabilities range from weak authentication and SQL injections to a lack of verification, encryption and adoption of software updates. A survey concluded in 2016 that consumers are increasingly aware of these risks and postpone the adoption of IoT as the risks currently outweigh the potential added value [8].

Lastly and perhaps most importantly, the adoption of the IoT is slowed by a great absence of business planning and models in many IoT projects. Due to a disproportionate focus on the technological aspects of these projects, business model innovation is mostly lacking [32]. Creating innovative business models could usher most sectors into the Fourth Industrial Revolution. Well defined value propositions will outline the potential benefits from a correct adoption of IoT, not only for business but also for consumers. By having a clear understanding what IoT offers, businesses can better predict the financial impact of IoT projects. This enables improved project planning and focus on important implementation aspects. These aspects include, among others, user experience, customer value and the previously described challenges of standardisation, and privacy and security.

1.2 IoT in Transportation

One of the prevalent sectors for applying IoT is transportation. While there is a variety in transportation systems, including land transportation, maritime transportation, aviation and more recent pipeline transportation, the word transportation is most often used to indicate land transportation. This can be divided further in rail and road transportation. In this paper, transportation will be used to indicate road transportation.

Transportation is, despite its major benefits, responsible for some of the most severe environmental and social conflicts. The most outspoken problem is that of air pollution and sustainability, which many claim is the main source of global warming [30]. The heightened demand of individual transportation puts immense stress on infrastructure, which limits are often stretched to a maximum.

IoT is said to offer major contributions in mitigating these problems. What ultimately simply entails the collection of more data, should lead to what some call Smart Mobility. Proper application of this data has added value for driver advisory systems, road maintenance and infrastructure planning, self-driving vehicles and traffic management and more [33]. This should ultimately lead to a more efficient use of transport infrastructure.

Modern road vehicles already collect large amounts of data, which are used by manufacturers and technicians for usability and repair analyses. For larger exploitation of vehicle data, the concepts of Intelligent Transportation Systems (ITS) [34], Vehicular Social Networks (VSN) and the Internet of Vehicles (IoV) [20] have been studied thoroughly. These studies, however, mainly focus on technical aspects as data collection [24][15], communication technology, cloud services [17], data analysis [10][21], etc. While this is incredibly important, little research has focused on enabling widespread exploitation of the collected and analysed data. The analysis of applicable business models that ensure the usage of this collected data by different stakeholders is severely lacking.

2. METHODOLOGY

In this paper different potential business models will be studied, that can maximise exploitation of vehicle collected data. Different stakeholders will be considered in finding compatible business models, ethical considerations, and supporting and enabling technology. These results will be fitted in a framework outlining the recommended strategies in various conditions.

To reach these goals, the following research questions will be answered:

RQ 1 Through what business models can the exploitation of IoT data collected from vehicles be maximised?

- **RQ 1.1** What characteristics does the collected data have?
- **RQ 1.2** Which stakeholders have a potential interest in the collected data, and why?
- **RQ 1.3** What conditions can be derived from the data and stakeholders that might warrant the use of different business models?
- **RQ 1.4** What business models fit these conditions best?
- **RQ 1.5** Can these business models be worked in an accessible framework to help business decision processes?

RQ 2 What technologies are necessary to correctly execute the business models and sustain stakeholders' interests?

This paper will consist of integrative literature research. This entails the reviewing and synthesising of representative literature on the topic of IoT in the transportation sector in an integrated way, such that new frameworks and perspectives on the topic are generated.

Through literature research, different types of data and stakeholders will be identified and listed. The data will be placed in different categories and possible use cases will be identified. The stakeholders will be evaluated for their different use cases and possible requirements of data use and acquisition. These results will be cross referenced to distinguish topics that are of influence when selecting a business model. Further literature research will identify business models that can be used to exploit the collected data. Their compatibility with the found requirements will be evaluated. This evaluation will result in a framework that might be used in the decision-making process for exploiting IoT collected vehicle data.

3. COLLECTED DATA

Arguably the largest collector of road vehicle data is the automotive industry. This sector has been relying on the collection of performance and usage data for a long time. With the introduction of electric and self driving cars, the big car data market could reach 750 billion dollars in revenue by 2030, according to McKinsey & Company [7]. In this same report, McKinsey identified five macro-categories of data ordered by the perceived privacy sensitivity:

- **External road and environmental conditions:** This could include types of camera and sensor data that register the environment of the car, ice or fog warnings for example.
- **Technical status of the vehicle:** This could include data such as oil temperature,
- **Vehicle usage:** Includes data such as speed, location and load weight
- **Personal data and preferences:** This contains data such as passenger identities, radio use, application use, etc.
- **Direct communication from the vehicle** This pertains communication data such as calendar, telephone, SMS and e-mail.

These categories identify types of data that could be available, but not whether this data is currently being collected by manufacturers. To ascertain if this is the case, different data collection practises have been inventoried and displayed in Appendix A. For this overview, privacy policies and terms of service of three car manufacturers [28][12][9] have been used to identify which data they collect. Manufacturers of other vehicles have not been included, as they did not advance as far as car manufacturers and less (or similar) data is generated by these vehicles.

From this inventory can be concluded that all data currently being collected fits within the five macro-categories, with no data put under 'other'. No personal communication data is being collected, by this set of manufacturers. Vehicle status and usage, and personal data and preferences seem to contain the bulk of the data, making most of the collected data privacy sensitive. This is due to the conclusions that can be made about an individual's behaviour in the car that may be of interests to insurance companies and authorities. Environmental data is also collected, primarily for R&D or providing services. Furthermore, most data seem to be collected continuously, with BMW not sharing their collection methods.

This inventory does not include data that might be collected in the future. Additional data might become available in future cars, such as bio-metric data and in-car shopping information, which creates a larger set of personal data. To accommodate this data, the fourth category can be split in 'Personal data' and 'Personal preferences'.

3.1 Data profiles

This leaves a total of six macro-categories that together enable the ability to create a full customer profile by asking the following questions.

- **External road and environmental conditions:** What are the customers' environment and context?
- **Technical status of the vehicle:** How responsible is the customer for his/her property?
- **Vehicle usage:** How does the customer behave?
- **Personal preferences:** What does the customer like?
- **Personal data:** Who is the customer?
- **Direct communication from the vehicle:** How does the customer think?

A similar profile can be made about the product and services themselves, by asking the following questions.

- **External road and environmental conditions:** What are the environment and context in which the product or service is used?
- **Technical status of the vehicle:** What is the status of the product or service during customer interaction?
- **Vehicle usage:** How is the product or service being used?
- **Personal preferences:** What preferences does the user have for the product or service?
- **Personal data:** Who uses the product or service?
- **Direct communication from the vehicle:** What products and services are consumers thinking about?

3.2 Data aggregation

By aggregating data from individual customers, macro-economic information becomes available. The validity of this information increases with the availability of aggregation data. This means data from single manufacturers would not suffice, warranting the combination of all available data from multiple manufacturers and infrastructural measurements. This aggregation of data then creates a macro-economic overview based on the following questions.

- **External road and environmental conditions:** What is the status of infrastructure (e.g. road and building quality)? What is the status of the environment (e.g. air quality, weather conditions)?
- **Technical status of the vehicle:** Where is traffic congestion likely to occur? What vehicle maintenance, repair and sales can be expected?
- **Vehicle usage:** Where is traffic congestion likely to occur? What patterns (e.g. travel times, traffic flows) occur in traffic? What financial risks are expected (for insurance and lease companies)?
- **Personal preferences:** What preferences does the user have for the product or service?
- **Personal data:** Who uses the product or service?
- **Direct communication from the vehicle:** What products and services are consumers thinking about?

3.3 Data Ownership

One important consideration that needs to be made, is who the owner is of the collected data. In a business context, the data owner is the person or group who is responsible for data assets within the business. In this paper, data ownership is understood as the set of exclusive rights and control over the data, similar to ownership in the natural world. In this case, the decision needs to be made whether the collected data should be owned by the device owner (consumer, lease company, etc.) or the device manufacturer.

This is important, as property is generally understood as being self-propagating: the owner of property also owns any economic benefits derived from said property. According to this, the data gathered by IoT devices, in this case vehicles, should be owned by the owner of that vehicle. This is opposing current practises in the IoT space. Generally, IoT data collected from consumer devices is send to and exploited by the manufacturer of said device. This possible because data ownership has not been explicitly defined in legislature, with copyright laws often falling short [25]. The most advanced privacy legislation, the EU's GDPR [23], explicitly does not speak of data owners, but gives rights to data subject to protect their privacy. This means that businesses can lay claim to data collected from the devices they produced. The GDPR requires businesses to base their claim on given requirements, one of which is consent. But this only holds for personal data, not for the non-personal data which is being collected from consumer owned IoT devices.

3.4 Privacy

Part of the data ownership discussion should be the privacy of the data subject. Here the question should be asked, how much does the collection of this data impact the data subjects' privacy. McKinsey recognised this by adding a spectrum of privacy sensitivity to their macro-categories.

There are three types of data in this regard: anonymous, pseudonymous and personal data. Pseudonymous data gained a renewed interest with the introduction of the GDPR, as it helps to comply with the secure data storage requirements. Pseudonymization involves the removal of any direct identifiers of the data subject. As such, the data can still be related to the data subject when analysed for behavioural patterns and combined with other data. This could be, for example, when patient-related data has to be passed on securely between clinical centres. Due to the reversible nature of pseudonymized data, it still falls under GDPR rule [29].

The raw data collected from connected vehicles should be qualified as personal data. This because the data holds information about the data subjects' preferences, behaviour and movements. Secondly, the data communication holds identifiers of the communication device, which is often integrated in the vehicle. With these and other identifiers stripped away, most data becomes pseudonymized data with the exception of direct communication data. Direct communication data holds names, dates, locations, etc. all on the data itself. Pseudonymizing this would mean severely compromising the integrity of the data. Further anonymizing the data would require the data set to be divided and scattered throughout the data space, such that the original set can never be recombined.

Depending on the specific collection and storage methods of the manufacturer, the collected data can be qualified differently. First, it is important to decide to what extend the collected data should be processed and anonymized. Second, it is important that the stored data receives some privacy sensitivity qualification, such that it influences business decisions about further exploitation and compliance with the GDPR (or other legislature).

4. STAKEHOLDERS

Historically, the value from a product is derived along the products value chain. The concept of the value chain was first introduced by Micheal Porter and has since been used as a support tool in business strategy decisions. Porter introduced Inbound Logistics, Operations, Outbound Logistics, Marketing and Sales, and Service as primary activities, and Procurement, Human Resource management, Technological Development and Infrastructure as secondary activities [22]. On an industry level, this is better known as the supply chain, where external parties are introduced each with their own value chain. Traditionally, this supply chain is unidirectional as it is focused on the development of a product. As the connection between parties in the supply chain is pre-existing, this can be leveraged to identify the first set of stakeholders that are interested in the collected data. By introducing this data, the supply chain becomes omnidirectional as information and therefore profits now interconnect every stakeholder. This enables new chain structures and business opportunities.

4.1 Automotive Supply Chain

Appendix B shows a set of parties that are currently part of the automotive industries supply chain from the manufacturers point of view. It also outlines the different data interests of these parties, and challenges they face in exploiting this data together with solution strategies [4]. Some fields are empty, as this information was not provided by the used literature. Nonetheless, they might still have an interest in the collected data and most likely face challenges that are similar to the other parties.

Though the roles differ between parties, most seem to have similar data interests. They all revolve around improving R&D, predictive maintenance, providing individualised products and services and improving customer interaction. A similar trend can be seen in the challenges they face. Technical challenges are predominantly data management, human-machine interfacing and identifying individuals. Business challenges are understanding and communicating the value proposition and aligning interests and vision internally and externally. Most of these challenges require more knowledge about data governance and opportunities, and better communication within collaborations.

4.2 External and New Stakeholders

Of course, there are more parties that derive value from the automotive industry. These parties, however, are not directly part of the industries supply chain. With the growing availability of vehicle data, new stakeholders are also introduced. Appendix C shows an overview of these parties and their data interests. These parties have been identified by multiple researchers [4][7][5] and a venture that aims to a create car data exchange platforms [3][2]. Though they did not identify possible challenges, it can be assumed that there are similar to the previous set.

Here many different data interests arise. This is mainly due to the different characteristics of the parties. Some of these differences are business versus consumer focused, service versus product provider and profit versus non-profit. Though the specific use cases differ, there are some commonalities visible. First is more informed decision making and improved R&D, which are used for creating a long-term business strategy. Second is individualised offers and advertisements, which are derived from more personal information. Last is better resource allocation and predictive maintenance, based on expected use. This is similar to the stakeholders that are already part of the automotive supply chain.

4.3 Stakeholders and Data

Different data types are needed to support these interests. Similarly, each use case might require different data streams, be it continuously, in intervals or just once. Appendix D creates an insight in these relations, by evaluation which macro-categories are needed for which use case and on what interval.

This table shows that the first three macro-categories, External road and environmental conditions, Technical status of the vehicle and Vehicle usage, are mainly used for use cases that are product focused or help with creating long term strategies. The last three macro-categories, Personal preferences, Personal data and Direct communications from the vehicle, are required for customer focused and short-term strategy use cases. It is important to note that some use cases require the aggregation of data from several macro-categories. This means that, while data can (and should) be categorised, complete separation is not desirable.

Furthermore, most use cases require data to be received periodically. The minimal period length needed is not given, but can vary between hourly, daily, weekly or yearly. In only two cases a continuous data steam is required. This goes for use cases where continuous monitoring or connectivity is required (e.g. use-based insurance and IT service providers). Lastly there are two cases which only require data 'on-call', These are cases where actions depend on instigation from the consumer (roadside assistance and product delivery). Appendix D does not display the need for one-time data availability for the given use cases. To prevent ambiguity in the table, one-time data availability was seen as part of the periodic data need. It is entirely possible that start-ups, for example, need a single static data set to test and develop their product or service on. If this proves successful, more true periodic data steams might be required.

5. BUSINESS MODELS

Monetization is defined as the act of making money from something. In this case, it is the exploitation of data to create new revenue streams. The strategy of a business for monetization is a business model, most importantly containing a value proposition. Big data business models differentiate from traditional business models, in that it concerns a non-rivalrous product or service that has a wide applicability range. Added complexity originates from the often private nature of the data, proving data ownership and authenticity and preventing duplication and misuse.

McKinsey identified three macro-categories of value-creation models [6], some of which already became apparent in the Stakeholders section.

- **Generating revenues**

- **Direct monetization:** Selling products, features, or services to the customer
  - **Tailored advertising:** Leveraging car data to push individual offerings to customers
  - **Selling data:** Collecting, analysing, and reselling big data to third parties
- **Reducing costs**
  - **R&D and material costs reduction:** Gathering product field data for development
  - **Customers' costs reduction:** Analysing actual usage patterns to reduce repair and downtime costs
  - **Improved customer satisfaction:** Better tailoring product/services to customer needs
- **Increasing safety and security**
  - **Reducing time for intervention:** Collecting and forwarding warnings in real time, pointing in the right direction

Most of these models concern 'internal' monetization models: creating value within the organisation. While these internal models can be difficult to concretize, they are less challenging to implement than external monetization models since they require little external collaboration and face less privacy and security related challenges.

5.1 External Monetization

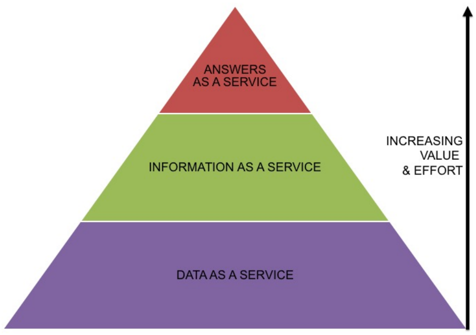

Figure 1. Three core big data business models and the value they bring [1]

Data can be monetised externally by 'selling' data. Generally, three different business models can be utilised, each increasing in value but also effort [1]:

- **Data as a Service:** This model has a value proposition that allows the data customer to employ their own analytic tools to derive conclusions from the data. In this case near raw data is offered. This warrants thorough pre-emptive processing of the data to erase any sensitive or personal information.
- **Information as a Service:** This model required the analysis of data to derive information. This information is then part of the value proposition and can be used by the data customer to derive answers. Data customers might not have the knowledge or tools to analyse data but are willing to exchange value for this analysis. Main actions required for this model are therefore data analysis and visualisation.
- **Answers as a Service:** This model is focused on providing high-level answers to specific customer questions. Here, there is very little actions required from the data customer side, resulting in more given value. It does require an extensive understanding of the data customers needs however, which might prove difficult.

Gandhi et al. [13] introduce similar models: Data as a service, Insight as a Service and Analytics-enabled platform as a service. The last is described as using *'sophisticated and proprietary algorithms to generate enriched, highly transformed, customized real-time data delivered to customers via cloud-based, self-service platforms.'*

Summarising, data can be monetised by selling (providing access to) near-raw data, selling information derived from analytics and selling answers to concrete customer needs. What data, information or answers are required highly depend on the customers use case and needs to be identified individually.

5.1.1 Pricing

Pricing has always been a difficult process, and this is not different for data. In fact, it can be argued that putting a price on data is a more difficult task. This can be caused by unclear value propositions and operational risks. Pricing is nevertheless important, as it has an impact on the revenue generation. There are two main ways to price data products [18]:

- **Cost-based pricing:** Based on the costs made to create the data product plus a margin. Costs include set-up, personnel, overhead and on-going IT costs.

  It is recommended that an overview of the costs is created to prevent losses, but not to set a selling price. Cost-based pricing should be regarded as the minimum price. This is due to the relatively low fixed and variable costs, resulting in a margin which is much higher than that in other cost-based industries.
- **Value-based pricing:** Based on the value the customer will derive from the data and, consequently, what they are willing to pay. This way, the majority of the available value is captured. It is, however, rather difficult to find a balanced price point and requires constant consideration and self-assessment.

  Factors that might influence the price include brand strength, data collection frequency, completeness, volume, rarity, organization and reliability of the data set and the ease of using it.

5.1.2 Data Marketplaces

Data marketplaces are not different from normal marketplaces, in that it is a platform where buyers and sellers connect. In practice, it allows data owners and data consumers to sell and buy data through graphical (file upload/download) or back-end interfaces (API connection). There are different types of marketplaces for different kinds of data. They can, for example, be divided in marketplaces for personal data, business data and IoT data [26].

- **Personal data marketplaces:** Aims to allow consumers to monetize their data directly and thereby controlling their privacy.
- **Business data marketplaces:** Aims to support efficient B2B exchanges of structured and large sets of data.

- **IoT data marketplaces:** Aims to allow for the purchase of real-time data feeds of remote devices for a price per-period or per-request

Due to the platformed nature of data marketplaces, they allow for crowdsourcing and standardisation of data, and fairness between sellers who set their own prices and buyers who choose from whom they buy. Thanks to the trust and security enabled by Blockchain technology, data marketplaces are said to enable data flows achieving a total value of 3.6 Trillion dollar and over 4.4 zettabytes transferred yearly by 2030[27].

Data marketplaces can be either private or public. Public marketplaces are often developed by third parties and offer wider connectivity between different types of data sellers and buyers. Therefore, every participant is dependent on the infrastructure and functionality offered by that third party. Most public marketplaces are based on blockchain technology, voiding the need for verification and trust in the marketplace owner. Private marketplaces are often developed by big data sellers themselves, thus maintaining more control over infrastructure and functionality. It does, however, require dedicated attention to facilitate this service, possibly limiting revenue streams. As most vehicle manufacturers seem to miss know-how of data monetization strategies and technology, finding a well-fitting public marketplace is recommended. This minimises the required effort and ensures wider data standardisation. Possible partnerships with such marketplaces can be set up, to introduces new and needed functionality that supports all stakeholders.

5.2 Data, Stakeholders and Monetization

When it comes to external monetization of data, three different models can be utilised through three different marketplaces: DaaS, IaaS and AaaS, and personal data, business data and IoT data marketplaces.

IoT data marketplaces are specifically designed to facilitate continuous or high-frequency periodic data streams. They are also focused on providing (near) raw data directly from IoT devices, fitting with the Data as a Service model. While this platform can be used for every macro-category of data, the Data as a Service model seems to only fit to a limited set of use cases. Only the use cases of High-tech giants, start-ups and entrepreneurs seem to require raw data. This means that, while this business model and marketplace is important, it does not support the bulk of the data value creation. The sharing of raw data also holds major privacy implications, as this continuous and real-time data steam can easily be related to the vehicle owner.

Personal data marketplaces support the monetization of data that describes the characteristics of a person or is otherwise privacy sensitive. This data is considerably less variable than raw IoT data. This does not mean that there can not be overlap between the IoT data marketplace and the Personal data marketplace. Location data, for example, can be used in both marketplaces. The main difference is that the IoT data marketplaces would live-stream this data, while the personal data marketplace accumulates this in an overview about the individual's movements. All business models can be used, depending on the functionality a specific marketplace offers. Ideally, the Data as a Service model would be unnecessary, due to well implemented support for the Information as a Service or the Answers as a Service model by the marketplace. It is required that the marketplace offers these tools for use cases with individual data subjects or multiple data subjects. This is because individual data subjects often do not have the know-how of data analytic tools to provide IaaS or AaaS themselves, or do not have access to data from multiple subjects. By providing these tools, data subject privacy is ensured since direct raw data access is no longer required. It also means that the data buyer no longer has to acquire and execute the data analytic tools themselves. If proprietary analysis of the data is required, stakeholders could be allowed to run their analytic tools on the marketplace without direct access to the raw data, enabling a more privacy secure Data as a Service model. Assuming that these functionalities are provided, most use cases are supported by this marketplace.

Similar to personal data marketplaces, business data marketplaces could support all business models. Here, the privacy aspect should be considerably smaller, as data should be anonymized or consent should be given by the data subject. When utilising the IaaS or AaaS model, this should be minimised even further. Again, most use cases would be served through this marketplace, assuming given consent for the sharing of personal data. Contrary to personal data marketplaces, business data marketplaces are not required to offer data analytic capabilities on the platform. Businesses have more capabilities to provide data analytic services themselves. Providing this service could create an extra revenue stream and potentially even new competitive advantages. It also ensures closer collaboration in traditional and new value chains.

The key difference between the personal data marketplace and the business data marketplace is data ownership. If the consumer is given ownership over the collected data, he/she should be allowed to exploit their data through a personal data marketplace. Otherwise, the corporation owning the data can use the business data marketplace. The IoT data marketplace can be used in both cases and potentially in parallel with the other marketplaces.

6. DATA BUSINESS MODEL FRAMEWORK

There are several factors important in the data monetization decision making process. This starts by determining the data owner. The data owner decision is based on a weighing of customer and business interests, as legislation gives little direction. This weighing may be influenced by the privacy sensitivity of the data. Here the type of data is not of direct importance. This is due to the high variability of data types, which requires global overarching strategies and decision as opposed to specialised strategies per data type. In this case, the data type must be used to determine the privacy sensitivity to the data subject.

Given that the raw collected data constitutes personal data, either consent must be acquired or, pseudonymization or anonymization steps need to be taken in order to sell the data. This is again mostly based on a weighing of customer and business interests, although in Europe GDPR compliance must be met.

The second important factor is stakeholder interests. Here, again, specific use cases are not of direct importance, but must be used to determine what service the stakeholders most require. In most cases, some form of data analytic is required by the stakeholders. For this Data as a Service and Information as a Service can be utilised. Data as a Service can only be used with consent from the data subject or anonymization of the data. Answers as a Service can also be provided, but this requires extensive knowledge of the specific stakeholders and the data analysis tools to

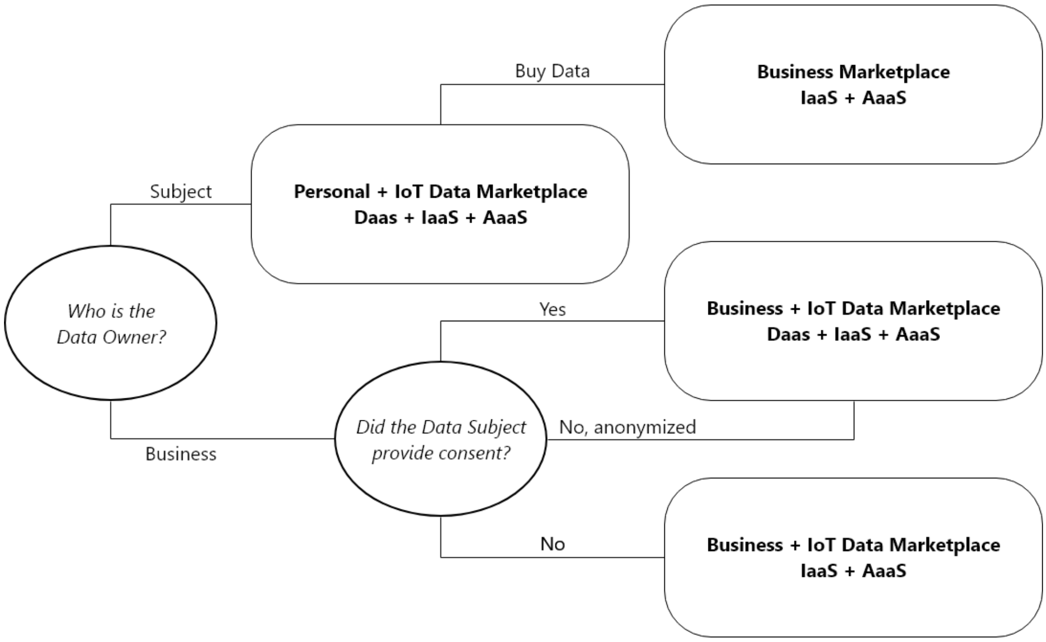

Figure 2. Data Business Model Framework

be used. It is important to state that these different services are not mutually exclusive and can thus be offered simultaneously.

Third important factor is the type of marketplace to be used. This is mostly dependent on who is deemed the data owner. Furthermore, a choice can be made between using a private or public data marketplace. The latter is mostly preferred, as it requires less knowledge of data analytics and little to no entry investment.

These factors resulted in the Data Business Model Framework as shown in Figure 2. This framework contains a decision tree which can be used during the first part of the data monetization decision making process.

The tree starts with the Data Owner decision. In case the data subject is selected as the data owner, connected between the connected vehicle and personal and IoT data marketplaces should be provided. In this case, data can be bought from the data subject and used to provide IaaS and AaaS through a business data marketplace. If the manufacturer is deemed data owner, consent needs to be acquired to sell personal data.

If consent is given, business and IoT data marketplaces can be used to provide DaaS, IaaS and AaaS. If consent is not given, anonymization can be used to offer the same services. If the data is not anonymized Data as a Service can no longer be offered.

The second part of the decision-making process involves shaping the different services and selecting fitting data marketplaces. This, however, is a highly complex process which is highly dependent on the use cases of different stakeholders. Stakeholders might require the ability to run proprietary algorithm for their data analytics. This is difficult, as these algorithms are confidential and can not be shared with the data owner for analysis. But allowing stakeholders to run their algorithm with access to raw data no longer ensures data privacy and security. A solution could be to find a marketplace in which stakeholders are enabled to run their algorithms on a provided platform, where data owners can verify the results of that analysis. Therefore, manufacturers should create an overview of possible use cases and design their Information as a Service model to fit those use cases. Therefrom, different requirements can be derived that can be used in selecting a public data marketplace.

Data as a Service and Answers as a Service have little influence on the data marketplace selection. This is because Data as a Service simply requires a data steam and payment system, and Answers as a Service can be handled similarly if analysis is done internally and the results offered on the marketplace.

Lastly, some global requirements need to be set up. These can be the pricing and payment structure of the marketplace, compatibility with the industry, number of available buyers, etcetera.

7. CONCLUSION

The connected vehicle data monetization space is extremely diverse. This start with the wide variety of possible collected data and different decisions made by businesses regarding privacy and data ownership. Added to the complexity is the wide variety of stake holders and use cases, which require a near infinite amount of different data analytic executions. This paper introduces a basic framework to shape the beginning of the data monetization decision-making process and guidelines for completing this process.

Part of these guidelines is the recommendation to utilise public data marketplaces fitted best to the situational requirements. Many of these marketplaces, however, are in a very early stage of development and thus offering limited functionality. This creates an opportunity for manufacturers as large data sellers for collaboration with these marketplaces to implement the required features.

Further research should unveil how manufacturers can derive the requirements based on which data services can be given shape and data marketplaces can be selected. This can than form an addition to the basis proposed in this paper. Furthermore, this framework has to be tested against current industry practises to ensure viability and usability.

Data is by many regarded as the new oil and connected vehicles have the ability to collect a varied set of data types. The sharing and monetization of this data creates entirely new value steams for manufacturers and might even re-define economies worldwide. As such, businesses should utilise data marketplaces and sharing platforms to ensure the widest possible utilisation of the available data. This benefits both the business itself and society as a whole.

8. REFERENCES

[1] Exploring Big Data Business Models - Business Models Inc.

[2] HERE Technologies | HERE.

[3] Otonomo | Automotive Data Services Platform for Connected Cars.

[4] S. Aggarwal, J. Deichmann, R. Dhawan, P. Gao, and F. Hansson. From Buzz to Bucks - Automotive Players on the Highway to Car Data Monetization. *McKinsey&Company Center for Future Mobility*, March 2018.

[5] A. Andersson and M. Axelsson. Creating and Appropriating Value from Connected Vehicle Data. 2018.

[6] J. Balasubramanian, S. Beiker, S. Chauhan, T. Colombo, F. Hansson, and S. Inampudi. Car data: paving the way to value-creating mobility - Perspectives on a new automotive business model. *McKinsey & Company Advanced Industries*, (March), 2016.

[7] M. Bertoncello, G. Camplone, P. Gao, H.-W. Kaas, D. Mohr, T. Möller, and D. Wee. Monetizing car dataâĂŤnew service business opportunities to create new customer benefits. *McKinsey & Company*, 2016.

[8] A. Björnsjö, M. Viglino, and G. Lovati. Igniting growth in consumer technology. *Accenture Report*, 2016.

[9] BMW Group Nederland. Juridische informatie over gegevensbescherming, 2019. [Online; accessed 19-May-2019].

[10] N.-E. El Faouzi, H. Leung, and A. Kurian. Data fusion in intelligent transportation systems: Progress and challenges–a survey. *Information Fusion*, 12(1):4–10, 2011.

[11] R. Fantacci, T. Pecorella, R. Viti, and C. Carlini. Short paper: Overcoming iot fragmentation through standard gateway architecture. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*, pages 181–182. IEEE, 2014.

[12] Ford Nederland B.V. Connected cars privacyverklaring, 2019. [Online; accessed 19-May-2019].

[13] S. Gandhi, B. Thota, R. Kuchembuck, and J. Swartz. Demystifying data monetization, 2018.

[14] S. Ganguli and T. Friedman. Iot technology disruptions: A gartner trend insight report, June 2017. [Online; accessed 07-May-2019].

[15] L. Guo, M. Dong, K. Ota, Q. Li, T. Ye, J. Wu, and J. Li. A secure mechanism for big data collection in large scale internet of vehicle. *IEEE Internet of Things Journal*, 4(2):601–610, 2017.

[16] S. Haller, S. Karnouskos, and C. Schroth. The internet of things in an enterprise context. In *Future Internet Symposium*, pages 14–28. Springer, 2008.

[17] W. He, G. Yan, and L. Da Xu. Developing vehicular data cloud services in the iot environment. *IEEE Transactions on Industrial Informatics*, 10(2):1587–1595, 2014.

[18] E. Leonidas and W. Walker. A Practical Guide to Pricing Data Products.

[19] A. Mashhadi, F. Kawsar, and U. G. Acer. Human data interaction in iot: The ownership aspect. In *2014 IEEE world forum on Internet of Things (WF-IoT)*, pages 159–162. IEEE, 2014.

[20] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu. Vehicular social networks: Enabling smart mobility. *IEEE Communications Magazine*, 55(5):16–55, 2017.

[21] B. Pan, U. Demiryurek, and C. Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *2012 IEEE 12th International Conference on Data Mining*, pages 595–604. IEEE, 2012.

[22] M. E. Porter. *Competitive advantage: Creating and sustaining superior performance*. Simon and Schuster, 2008.

[23] G. D. P. Regulation. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. *Official Journal of the European Union (OJ)*, 59(1-88):294, 2016.

[24] M. Reininger, S. Miller, Y. Zhuang, and J. Cappos. A first look at vehicle data collection via smartphone sensors. In *2015 IEEE Sensors Applications Symposium (SAS)*, pages 1–6. IEEE, 2015.

[25] T. Scassa. Data Ownership. *SSRN Electronic Journal*, sep 2018.

[26] J. Smith. Data Marketplaces: The Holy Grail of our Information Age, 2018.

[27] D. Tang, J. Fujii-Hwang, E. Colwill, S. Arora, A. Shah, A. Mendizabal, and R. CallejasâĂŃ. Value of Data: The Dawn of the Data Marketplace | Accenture, 2018.

[28] Tesla Inc. Privacy & legal, 2019. [Online; accessed 19-May-2019].

[29] V. Villers, F. Vonner, and C. Nédélec. Anonymisation and pseudonymisation. *PricewaterhouseCoopers Legal LLP*, 2016.

[30] M. P. Walsh. Highway vehicle activity trends and their implications for global warming: the united states in an international context. *Transportation and Global Climate Change*, 1993.

[31] M. Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–105, 1991.

[32] M. Westerlund, S. Leminen, and M. Rajahonka. Designing business models for the internet of things. *Technology Innovation Management Review*, pages 5–14, 2014.

[33] X.-F. Xie. Key applications of the smart IoT to transform transportation systems. Technical Report WIO-TR-18-001, WIOMAX, 2018. [Online; accessed 19-May-2019].

[34] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639, 2011.

APPENDIX

A. DATA COLLECTION TABLE

Manufacturers		Macrocategories						Used Categories
		External road and environmental conditions	*Technical status of the vehicle*	*Vehicle usage*	*Personal data and preferences*	*Direct communications from the vehicle*	*Other*	
Tesla	*Collected data*	Camera images Video clips of accidents Sensor data of accidents Road segment data Short video clips of external cameras	Electrical system functions Software version information Infotainment system data Safety-related data Data report to confirm that your vehicle is online Information about the current software version Vehicle connectivity information Data about any issues that could materially impair operation of the vehicle Data about any safety-critical issues Data about each software and firmware update	Vehicle Identification Number Speed information Odometer readings Battery use management information Battery charging history Braking and acceleration Security e-brake Information regarding the use and operation of Autopilot, Summon, and other features Current location of the vehicle Data about accidents involving the vehicle Data about remote services Charge rate and charging stations used	Personal settings in the vehicle Contacts Browsing history Navigation history Radio listening history Customer name Repair history Outstanding recalls Bills due Customer complaints	-	-	Telematics log data Remote analysis data Other vehicle data Service history Charging station information Advances features
	Collection method	In person Via remote access	In person Via remote access	In person Via remote access Via dynamic connection	Via dynamic connection	-	-	
	Collection goal	To improve vehicles and services towards the customers To be used by features in the vehicle To develop and improve autonomous safety features	To improve vehicles and services towards the customer To help improve their products and services	To improve vehicles and services towards the customer To diagnose and resolve issues with the vehicle To help improve their products and services To analyze which charging stations are being utilized, how long and efficient battery charges are, and where additional charging stations are needed.	To diagnose and resolve issues with the vehicle To facilitate the servicing of your car	-	-	
Ford	*Collected data*	-	Hardware model- and part numbers Status of vehicle systems Information on the performance of the vehicle Connection to Wi-Fi and mobile networks Software- and firmware versions of vehicle systems	Vehicle Identification Number Odometer readings Speed Use of gas pedal, brakes, steering wheel, seatbelts and other information about the use of the vehicle Current location of the vehicle Driving direction	Operation and use of systems and vehiclecabintechnologie like; navigation, infotainmentssystems, telephone-/Bluetooth connections, climate control, chair positioning, SYNC, changes in connection settings	Identification information SIM and electronic Serie number	-	Vehicle information Drive characteristics and -behaviour Information about services in the vehicle Information about the exact location Notifications about connection status Information about over-the-air updates
	Collection method	-	Via remote access using FordPass Connect, AppLink or Wi-Fi In person	Via remote access using FordPass Connect, AppLink or Wi-Fi In person	Via remote access using FordPass Connect, AppLink or Wi-Fi In person	Via remote access using FordPass	-	
	Collection goal	-	To improve connectivity services For diagnosing and resolving issues For product research and improvement To provide personalised products and services	To improve connectivity services For diagnosing and resolving issues For product research and improvement To provide personalised products and services To offer services for which the information is needed, such as discount on insurance	To improve connectivity services For diagnosing and resolving issues For product research and improvement To provide personalised products and services	To ensure FordPass Connect can send and receive data via the telecommunication network	-	
BMW	*Collected data*	Dynamic Traffic Information Environment information Sensor information Environment information about vehicle and charging location	Vehicle status data Vehicle Maintenance- and serviceproductdata Wallbox information Vehicle chargingmodus, planned depart times, charging settings, needed energy Information on charging points and assigned vehicles Information on energy prices Information on photovoltaic installation	Position- and movement data Chargehistoricdata Position- and movement data	User profile Address information of charging point	-	-	-
	Collection method	-	-	-	-	-	-	
	Collection goal	To provide ConnectedDrive contract services To provide Digital Charging Services	To provide ConnectedDrive contract services To provide Digital Charging Services	To provide ConnectedDrive contract services To provide Digital Charging Services	To provide ConnectedDrive contract services To provide Digital Charging Services	-	-	

Table 1. Data Collection by Vehicle Manufacturers

B. DATA INTERESTS SUPPLY CHAIN TABLE

Supply Chain	Stakeholder	Current role	Data Use Cases	Technical challenges	Business Challenges	Solutions
Inbound Logistics	*Raw materials*	Supplies raw materials to suppliers	Predictive Maintenance Warranty cost reduction Failure diagnostics R&D optimization	Data management	communicating value proposition limited understanding customer benefit Misaligned interests of isolated departments Further isolated from end consumer	Create clear use cases and implementation roadmaps Align interests with partners and start development collaborations
	Tier 2+ supplier	Supplies parts to higher tier suppliers				
	Tier 1 supplier	Supplies parts directly to manufacturer				
Operations	*Manufacturer*	Final assembly	Predictive Maintenance Warranty cost reduction Failure diagnostics R&D optimization Usage based insurance Advertising and delivering non-driving and driving related goods and services	HMI Data management Identification of and outreach to drivers	Communicating value proposition limited understanding customer benefit Misaligned interests of isolated departments make collaboration difficult Regulations surrounding data governance	Define clear vision of connected customer experience and share internally Redesign marketing Facilitate cross-functional collaboration and create hiring plan to close knowledge gap
Outbound logistics	*Warehousing*	Stores final product	-	-	-	-
	Transportation	Transports final product to, between and from warehouses	-	-	-	-
Marketing & Sales	*Marketing*	Often part of manufacturer	-	-	-	-
	Dealers	Sell final product	Improve customer interaction (e.g., explain features and offer services)	remote diagnostics in-vehicle advertisement driving style and user preferences	Gain better customer understanding	Collaborate with manufacturers and redefine business strategies Training sessions on selling connectivity features
	Retail	Sell components of final products				
Service	*Customer service*	Complaint resolution				
	Repair	Repair and maintenance of product				
	Third-party service providers	Provide services such as infotainment and navigation	Improve customer interaction Provide individualized services R&D optimization	No control of HMI Get data through manufacturer	communicating value proposition limited understanding customer benefit Manufacturers preferences Defining engagement strategy in car as an environment	Establish direct contact with customer Rapidly adapt to customer feedback Adapt to in-vehicle interfaces Redesign customer journey

Table 2. Data Interests Automotive Supply Chain Stakeholders

C. DATA INTERESTS OTHERS TABLE

Stakeholder	Current role	Data Use Cases
Insurers (Car, health, life, etc.)	Provide insurance through yearly contracts	Offer usage-based insurance Occasion-related policies Extended customer understanding
Roadside assistance	On-call repair service	Collect and process calls from vehicle sensors and automated alerts Optimize dispatching Analyse accident data for manufacturers and infrastructure operators
Infrastructure operators	Billing or toll roads Recharging/fuelling Road maintenance Parking services Traffic flow maintenance Urban planning	Optimize geographic deployment variable-pricing options Monitor asset status Communicate availability Predictive maintanance More accurate decision making
High-tech giants	-	Car data analytics services Offer IT backbone and front-end applications Develop in-car platforms and OS to boost data generation and provide seamless integration with other devices and environments
Start-ups	Smaller High-tech counterparts	provide innovative hardware/interfaces and monetization schemes
Mobility providers	Public transport Ride sharing e-hailing	Further improve public transport infrastructure Improve vehicle allocation, recharge and fleet operations
Retailers and service providers	-	Tailored in-vehicle and handheld advertisements Store location, opening hours and inventory optimization In-trunk goods delivery
Governments	Provide legistation regarding vehicle and road safety Set standard regarding data collection and sharing Create infrastructure strategy	provide data-enabled services (e.g. emergency call features) regulate technical certification, data ownership, IP rights minimize congestion and reduce accidents
Energy providers	Provide energy to vehicles	Location-based promotion On-demand services Usage insights and asset optimization EV infrastructure

Table 3. Data Interests External and New Stakeholders

D. DATA CATEGORIES STAKEHOLDERS TABLE

Stakeholders	Macro-categories						Frequency	Type
	External road and environmental conditions	*Technical status of the vehicle*	*Vehicle usage*	*Personal preferences*	*Personal data*	*Direct communications from the vehicle*		
Raw materials ***Tier 2+ supplier*** ***Tier 1 supplier***	Predictive Maintenance Warranty cost reduction Failure diagnostics R&D optimization			-	-	-	Periodically	Analytic results
Manufacturer	Predictive Maintenance Warranty cost reduction Failure diagnostics R&D optimization			-	-	-	Periodically	Analytic results
	Advertising and delivering non-driving and driving related goods and services							
	-	Usage based insurance		-	-	-	Continuous	Analytic results
Retail ***Customer service*** ***Repair***	-	-	-	Improve customer interaction (e.g., explain features and offer services)			Periodically	Analytic results
Third-party service providers	R&D optimization	-	-	-	-	-	Periodically	Analytic results
	-	-	-	Improve customer interaction				
	Provide individualized services							
Insurers (Car, health, life, etc.)	-	Offer usage-based insurance Occasion-related policies		-	-	-	Continuous	Analytic results
	-	Extended customer understanding					Periodically	Analytic results
Roadside assistance	Optimize dispatching Analyse accident data for manufacturers and infrastructure operators			-	-	-	On call	Analytic results
	-	Collect and process calls from vehicle sensors and automated alerts	-	-	-	-		
Infrastructure operators	Monitor asset status Predictive maintenance	-	-	-	-	-	Periodically	Analytic results
	-	-	Optimize geographic deployment	-	-	-		
	-	Communicate availability variable-pricing options		-	-	-		
	More accurate decision making			-	-	-		
High-tech giants	Car data analytics services Offer IT backbone and front-end applications						Continuous	Raw data
	-	Develop in-car platforms and OS to boost data generation and provide seamless integration with other devices and environments						
Start-ups ***Entrepreneurs***	provide innovative hardware interfaces and monetization schemes						Periodically	Raw data
Mobility providers	Further improve public transport infrastructure Improve vehicle allocation, recharge and fleet operations	-	Further improve public transport infrastructure Improve vehicle allocation, recharge and fleet operations	-	-	-	Periodically	Analytic results
Retailers and service providers	-	-	-	Tailored in-vehicle and handheld advertisements			Periodically	Analytic results
	Store location, opening hours and inventory optimization	-	Store location, opening hours and inventory optimization	-	-	-		
	-	-	In-trunk goods delivery	-	-	-	on call	Analytic results
Governments	provide data-enabled services (e.g. emergency call features) minimize congestion and reduce accidents			-	-	-	Periodically	Analytic results
	regulate technical certification, data ownership, IP rights							
Energy providers	Location-based promotion On-demand services Usage insights and asset optimization EV infrastructure			-	-	-	Periodically	Analytic results

Table 4. Data Categories Need of Stakeholders