A Business Model Analysis for Vehicle Generated Data as a Marketable Product or Service in the Automotive Industry

Frank Bodendorf and Joerg Franke
Friedrich-Alexander-University of Erlangen-Nuremberg (FAU)
Institute for Factory Automation and Production Systems
Erlangen, Germany
e-mail: frank.bodendorf@faps.fau.de
e-mail: joerg.franke@faps.fau.de

***Abstract*— Data-driven business models play a significant role in the digital transformation of traditional value-added industries. More and more existing and potential partners of automobile manufacturers show interest in the data generated by the vehicles. However, there is still no monetary value assessment to support decisions regarding the release of data. Traditional pricing approaches for material goods are based on cost, margin, and volume. However, these bottom-up calculation concepts are not applicable to digital goods. The background to this is, among other things, the uncertainty about the potential sales volume, the difficulty of cost splitting, and the high unit cost degression of digital goods. This paper provides a decision support for selling data to third parties as an intangible product. It introduces a concept that allows to value data generated by a motor vehicle in order to determine potential prospects and prices for sale. The evaluation model developed can be used to strengthen the car manufacturer's negotiating position towards potential data buyers.**

Keywords - Automotive Industry; Car Data; Business Model; Value Estimation

I. INTRODUCTION

The use of valuable data will fundamentally change competition in the future [31]. “The expected growth of the value pool from car data and shared mobility could add up to more than USD 1.5 trillion by 2030“ [1]. Volume and quality of this “data treasure“ will create strategic as well as operational competitive advantage [21].

Today, data is generated in large quantities by the vehicle, recording thousands of attributes. On the one hand, the vehicle user (driver) has the opportunity to enter data in on-board systems and “exchange” them for services. He/she is offered individually adapted functions, such as voice control, comfort settings when entering the car, navigational instructions in real time or other services [4][5]. On the other hand, a variety of sensors and computers in the vehicle, unnoticed by the driver, generates a steady stream of data, which among others serves for control purposes [14]. Examples are the anti-lock braking system or the automatic windscreen wiper and light regulation.

The data usage can be divided into nine purpose oriented categories [27]:

- Meeting regulatory and legal requirements (e.g., liability for material defects)
- Supporting marketing and advertisement (e.g., customer profiling)
- Assessing IT security (e.g., logging and monitoring)
- Improving technical processes (e.g., diagnostics and programming)
- Fulfilling terms of contract (e.g., new digital services and solutions)
- Innovating and developing products (e.g., monitoring and analytics)
- Ensuring road safety (e.g., traffic management)
- Transferring to third parties (e.g., car sharing)
- Facilitating vehicle use (e.g., autonomous driving)

All these categories have in common that value is created through the use of collected vehicle data. On the one hand, this value is reflected in technical or qualitative improvements as well as in cost reductions of the company's internal processes. On the other hand, the use of vehicle data can also lead to an economic improvement of the business results and in particular to an increase in turnover [32]. This may be a result of higher sales figures of products, i.e., manufactured vehicles, which are more attractive through data-based functions (“data infused products”). In addition, it is possible to offer certain data for sale as an end product itself [20].

This paper focuses on selling data to third parties. The demand for vehicle generated data depends on the benefit seen or expected by the buyer. From the perspective of the data provider, it is important to determine the value of the data in order to estimate the demand potential on an external market and to create appropriate pricing models [1].

II. EXISTING VALUATION APPROACHES

The generic value of a data product in sales situations cannot be determined by a benefit that has already been realized, since the data is not yet being used by the buyer at the time of the transfer. Therefore, an evaluation must be based on probable and potential benefits [20]. This value can be estimated by using qualitative and quantitative methodological approaches. For a corresponding systematic value determination of vehicle data, a number of existing evaluation methods are presented. This is initially done in a tabular overview in Section II.A, followed by a more detailed description in Sections II.B and II.C and by a discussion of limitations and transferability in Section III.A. Literature often speaks of data and information without exactly differentiating between these terms. Some authors see in information “refined data”, e.g., by placing it in a context of meaning. In this paper, both terms are used synonymously according to the quoted sources.

A. Overview of potential methods for data evaluation

The identification and selection of potential valuation procedures is done through a combination of literature review and in-depth interviews. First, 20 sources of literature are used to collect a comprehensive set of possible valuation approaches. Subsequently, valuation approaches are selected and specified with the help of 50 in-depth interviews with experts from the divisions or departments in the areas of cost engineering, data strategy, data analysis, and purchasing. The consolidated results are shown in Table I [27].

TABLE I. QUALITATIVE AND QUANTITATIVE VALUATION APPROACHES FOR DATA

Method	**Characteristics**			
	Type	*Input*	*Operator*	*Output*
Data Product Scorecard	Qualitative	Data attributes	Scoring-method	Willing-ness to pay
Data Value Design Canvas	Qualitative	Data use case	Expert workshop, Canvas nine factors	Inter-actions / connections
Value determi-nation per user	Quantitative	Acquisition cost Number of users	Discount calculation	Price per user dataset
Value improve-ment by data services	Quantitative	Data material	Statistical analysis, e. g, hypothesis testing	Increase in value or quality through the use of data
Value determi-nation by Laney	Quantitative	Data material	Gartner Valuation Model	Quali-tiative and financial value
Value determi-nation by partners	Quantitative	Theoretical value, maturity, expiration of information	Intangible Assets Evaluation	Monetary information value
Pricing based on customer value	Quantitative	Different data bundels	Versioning, price differen-tiation, surcharge calculation	Price for data bundels

B. Qualitative evaluation

The qualitative assessment and selection of methods is carried out by the procedure described in Section II.A. The following common methods for evaluating vehicle data are identified:

- Data Product Scorecard
- Data Value Design Canvas

The *Data Product Scorecard* is a method of pricing on data marketplaces. For this purpose, the customer's willingness to pay depending on various data properties must first be estimated. This qualitative evaluation of the data properties is made by the Data Product Scorecard from a simulated perspective of end users or potential buyers of the data [21]. As part of an evaluation workshop within the company, the role of the user is taken and each data characteristic given in the scorecard is rated with 0, 5 or 10 points.

The *Data Value Design Canvas* approach looks at the data value chain. The approach is based on the theory of Service Dominant Logic and the “Jobs-To-Be-Done” theory [3][22].

According to [25], the data value chain begins with the generation of data and extends up to the provision of information to the (paying) customer.

C. Quantitative evaluation

Many companies have problems finding the real economic value of their data [23]. For a rethink in the development of new business models [12] and the optimization of internal processes, the determination of this value, especially for the automotive industry, is of particular importance.

Value determination per user

When acquiring companies with data-driven business models who have not yet monetized their database but still offer data-based applications to the end user, the data value is often determined by the value of the application per user. The price of acquisition is divided by the total number of end users of the application. From this calculated price per user the average user acquisition costs are subtracted [16][28].

Value determination by Laney

According to Laney the data is evaluated through quality-based and quantitative financial analysis [20]. In the quality-oriented evaluation, the output is a scoring value between zero and one, in the financial evaluation it is an absolute monetary value. The two-part consideration focuses on methods for improving the “Information Management Discipline” and deals with “Foundational Measures” as:

- How correct, complete and exclusive is the data? (Intrinsic Value),
- How good and relevant is the data for specific purposes? (Business Value),
- How does this data affect key business drivers? (Performance Value).

On the other hand, the “Information Economic Benefit“ of “Financial Measures“ is examined:

- What would it cost us if we lose this data? (Cost Value),
- What could we get from selling or trading this data? (Market Value),

- How does this data contribute to our bottom line? (Economic Value).

Both considerations provide a quantitatively measurable contribution to the value of data and will be explained in more detail below. Based on a collection and analysis of existing valuation approaches according to Laney, a plausibility check and a requirement analysis lead to new valuation perspectives for the developed new process model (see Section III.A).

Basically, Laney's approaches are not limited to any specific field of application [33]. Section III.A discusses to what extent these approaches are suitable for vehicle data in the automotive industry and in what form adaptations and expansions are necessary for this.

Intrinsic valuation

The intrinsic value of information (IVI) is based on a consideration of information regarding quality and rarity.

$$IVI = Validity * Completeness * (1 - Scarcity) * Lifecycle \tag{1}$$

“Validity” reflects the proportion of “correct” data, “Completeness” the proportion of available data in all potentially accessible data, “Scarcity” the proportion of competitor data to the data available on the market and “Lifecycle“ the period of usefulness of the respective information.

Business value of information

When calculating the Business Value of Information (BVI), a process-specific value of data is aggregated for all corresponding processes of the customer’s company.

$$BVI = \sum_{p=1}^{n} (Relevance_p) * Validity * Completeness * Timeliness \tag{2}$$

BVI	Business Value of Information
n	Number of considered processes
p	Process index

“Relevance“ indicates the usefulness of the data or information for a business process. This value lies in the interval between zero and one. The probability measure “Timeliness” mirrors the aspect that data is not workable at every instant of time.

Performance value of information

The performance value of information (PVI) expresses how the performance of the object of consideration changes with the inclusion of data or information.

$$PVI = \left[\left(\frac{KPI_i}{KPI_c}\right) - 1\right] * \frac{T}{t} \tag{3}$$

PVI	Performance Value of Information
KPI_i	Key Performance Indicator using data
KPI_c	Key Performance Indicator without using data
T	Information lifetime
t	Period of KPI consideration

The key performance indices quantify the power of processes with and without data support. The quotient of both KPIs (KPI_i, KPI_c) conduce to the understanding how data can improve value creation. One example for the “performance value of information” in practice is the data driven Overall Equipment Effectiveness (OEE) for defect prevention [35].

Cost value of information

According to Laney, the cost of information is limited to the process costs for acquiring the data.

$$CVI = \frac{ProcExp * Attrib * T}{t} \left\{ + \sum_{p=0}^{n} LostRevenue_p \right\} \tag{4}$$

CVI	Cost Value of Information
$ProcExp$	Process expenditures
$Attrib$	Proportion of process costs for data acquisition
T	Average information lifetime
t	Period of the process cost measurement
p	Process index
n	Number of processes

The process expenditures “ProcExp” of an overall process are multiplied by a contribution factor “Attrib”. The contribution factor expresses which proportion of the process costs can be attributed to the acquisition of the data. “LostRevenue” takes into account the loss of revenue through information shortage in each process.

Market value of information

The market value of information (MVI) is relevant in public marketplaces. The market value of data is expressed, e. g., by a license price (exclusive price) multiplied by the number of licenses.

$$MVI = \frac{Exclusive\ Price * Number\ of\ Licenses}{Premium} \tag{5}$$

In addition to the market price and the number of licenses, the Premium factor takes into account the brand strength and rarity of the data. This Premium factor determines the extra charge a customer would be willing to pay to obtain exclusive rights.

Economic value of information

The Economic Value of Information (EVI) expresses the extent to which revenue changes as soon as the information is used to generate and increase revenue.

$$EVI = \left[\begin{matrix} Revenue_i - Revenue_c \\ -(AcqExp + AdmExp + AppExp) \end{matrix} \right] * \frac{T}{t} \quad (6)$$

EVI	Economic Value of Information
$Revenue_i$	Revenue with using the information
$Revenue_c$	Revenue without using the information
$AcqExp$	Cost for information gathering
$AdmExp$	Cost of information administration
$AppExp$	Cost for the information application
T	Information lifetime
t	Period of the measurement

First, the revenue is recorded without the use of certain relevant information. Subsequently, the turnover is estimated by a control group using this information. Also the expenditures for obtaining, storing, managing, and using information are identified. The respective revenues and expenditures are measured in a period t. The economic value of the information is now calculated by subtracting the revenues without information and the costs incurred from the revenues gained with the information. The resulting value is multiplied by the ratio of information lifetime T to duration of measurement t. This results in an EVI value for the full information lifecycle.

III. Developed Methodology

A. *Motivation of a new valuation approach*

IVI, BVI and PVI are pure comparative methods. These do not calculate monetary values, yet they show some interesting perspectives. The BVI sums up the value of special information over all use cases in the company, and both the rarity and the quality indicators are taken into account in the IVI.

The methodology introduced in this section follows an approach also based on PVI and BVI. The value of data can lead to process improvements and profit increases. The Data Value Design Canvas lists nine factors which affect the value of data. The effects of information/data are considered generally. For example, information/data protects against unwanted events or promotes wanted events. Unwanted events always result in costs. Thus, the avoidance of unwanted events corresponds to a cost reduction. The realization of desired events effects an increase in sales as the most relevant example.

The Data Product Scorecard assesses the willingness to pay of a customer. If the information considered is “perfect”, the customer is willing to pay the full price for this information. The determination of willingness to buy is based exclusively on estimates.

The method according to Laney, the Data Value Design Canvas and the Data Product Scorecard focuses on the quality of the information/data.

The monetary impact of EVI relates to an increase in sales. A monetary value can be derived from the Data Product Scorecard by multiplying the qualitative result score by the willingness to pay. A market value, in turn, may result from a fixed license price multiplied by the number of licenses sold or salable and a rarity and reputation factor.

All methods have in common that always certain use cases of data usage are considered. The PVI and BVI consider process improvements, the MVI and the Data Product Scorecard data sales, the Data Value Design Canvas both data sales and process improvements. Only the IVI does not aim at a defined application field because it is based on a general quality criterion.

By comparing the different approaches listed in Section II, some requirements can be derived for a new concept integrating different aspects:

- Quality factors must be taken into account.
- The willingness to pay is relevant for data sales. This depends on various factors, including the purchase motive, the perceived benefit, the reputation of the seller and the individual purchase situation.
- Data has the potential to increase sales on the customer side or to reduce costs for internal customer processes.
- Competition should be considered as an important factor.

B. *Evaluation model*

The developed “integrated methodology” for an innovative evaluation model meets these requirements by a combination of quality assessment [26], price differentiation [36], cost management, and competitive analysis [32]. So, for a specific use case, it is possible to estimate a monetary value of data by integrating the various value perspectives. The plausibility of selling prices is achieved by combining qualitative tools based on methods such as Business Model Canvas or Data Canvas with a practical evaluation through quality workshops as well as quantitative calculations. These include, among other things, the valuation by Laney, a bottom-up cost calculation as well as profit split approaches.

Figure 1 shows the process steps of the model for a use-case-specific value determination. The non-rivalry property of data enables multiple sales of similar data bundles or even the same dataset. The total value of the data bundle can be determined as the sum of the values across all (potential) use cases:

$$V_G = \sum_{i=1}^{n} V_i \quad (7)$$

V_G	Total value of data bundle G
V_i	Individual data value for the use case i
i	Use case index
n	Number of use cases

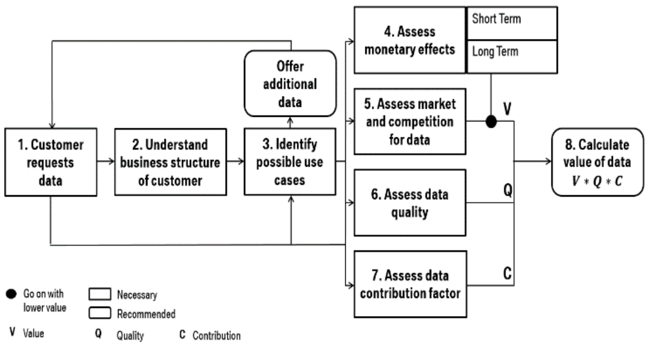

Figure 1. Process for the monetary valuation of data

As a first step in the process, data requests from potential business customers are collected or data is proactively offered to potential customers. In order to be able to identify potential data needs, a customer's business model must first be understood (1 + 2).

In which way the customer translates the data into benefits can be identified through a systematic analysis of possible use cases. For this a combination of the Business Model Canvas and Data Value Design Canvas is suitable. With progressive understanding of the application, it is theoretically possible to offer the customer additional versions of data bundles (2 + 3).

In order to determine the customer's willingness to pay, the value of the use case must be understood in detail [15]. For this, a possible cost reduction or increase in sales by the data must be determined.

For each use case, there is both a short-term and a long-term monetary benefit, which in individual cases can also be zero. The model of Figure 1 shows a parallel approach to Laney's business value calculation of information, which determines the data relevance for specific processes (4).

In cases where there is competition on "data marketplaces" or the self-collection of data is significantly more favorable than granting the monetary benefit to a third party, these influences must be measured for further calculation (5).

At the same time, the data is qualitatively evaluated based on selected criteria (see Section II.B). The model considers the monetary value of quality criteria, following Laney's valuation ideas in the form of a quality factor Q (see Section IV.D) (6).

In addition, the data contribution factor C takes into account that other vehicle generated data or additional information may be necessary in addition to the offered vehicle usage data (see Section IV.D) (7).

After having carried out steps 1 to 7, finally the use-case-specific combined value of the data is calculated. The determined preliminary data value V (see Section IV.B) is multiplied by the quality factor Q and the data contribution factor C. Both of them are discount factors. The preliminary value and the willingness to pay is reduced according to low quality or insufficient amount of data. Interdependencies between the factors are possible but not taken into account.

C. Restrictions

It is important to make the best selection of use cases to generate the maximum revenue of data sales for the business.

$$R = \sum_{i=1}^{n} RUC_i \tag{8}$$

R	Aggregated revenue
RUC_i	Revenue of use case i
i	Index of implemented use cases
n	Number of implemented use cases

The present technical restrictions must be observed. However, this limits the number of possible applications. The possible amount of data to be transferred to different use cases is restricted by the number of transmission paths and the transfer rate of each of them:

$$TL_k \geq \sum_{j=1}^{m} D_{jk} \tag{9}$$

TL_k	Maximum transfer rate for transmission k
D_{jk}	Data volume j to be transferred by path k
k	Index of possible transmission paths
j	Index of required data for each transmission path depending on the requirements of the application
m	Total number of required data volumes per transmission path

Since many use cases are based on fleet data, the complexity is additionally increased. Accordingly, it has to be determined which vehicles transmit which data via which path with which transfer rates, in order to technically enable the optimum amount of use cases. This problem cannot be solved conventionally or manually. It is necessary to test new tools for this, such as machine learning algorithms.

IV. Use Case Results

The applicability of the developed methodology is experimentally tested on the basis of three real sales situations. The first use case deals with the sale of weather data to a weather service provider who wants to offer additional hazard warning services for autonomous driving.

The second use case relates to the sale of weather data to a transmission system operator to “ensure a reliable and uninterrupted supply in the high voltage grid for approximately 41 million people” [34].

In the third use case the sale of road segment data (RSD) to a navigation maps provider is considered. This data helps to provide a high definition road map for autonomous driving.

A. Business structures of the use cases

Table II summarizes the results of the structural analysis of the business models. The findings serve as a first qualitative evaluation of the business structure, from which the data needs of the customers and the data bundles offered can be derived.

TABLE II. BUSINESS STRUCTURE ANALYSIS [17][18][29][34]

Object of investigation	**Customer**		
	Weather service	***Transmission system operator***	***Navigation maps provider***
Business Structure	***Value Proposition*** Deliver weather service ***Key resources*** Internal rational data: - External rational data: Satellite pictures Internal continuous data - External continuous data: Weather station information, weather car data ***Key Activities*** Gather and refine data for weather forecast ***Customers*** Companies, end users ***Segments*** Automotive, governments, software companies ***Relationships*** Direct ***Channels*** Direct contact ***Revenue Stream:*** Selling refined weather forecast information ***Cost Structure*** Data buying, personnel	***Value Proposition*** Secure reliable supply of electricity ***Key resources*** Internal rational data: Electricity prices External rational data: Consumer behavioral studies Internal continuous data Electricity supply, electricity demand External continuous data: Weather station data, weather car data ***Key Activities*** Transportation of electricity, maintaining energy balance ***Customers*** Companies ***Segments*** Large industries, consumers, governments ***Relationships*** Indirect contact ***Channels*** Customer events, customer committees ***Revenue Stream*** Offering connection and transmission services, maintenance of energy balance, operation of energy exchange, offshore balancing ***Cost Structure*** Grid connection, transmission services, maintenance of energy balance	***Value Proposition*** Build tomorrows road network ***Key resources*** Internal rational data: Street map data External rational data: - Internal continuous data Navigational data, destination External continuous data: Real time traffic information, car data (RTTI), Road segment data (RSD) ***Key Activities*** Gather and refine map data ***Customers*** Companies ***Segments*** Automotive industries ***Relationships*** Direct ***Channels*** Direct contact ***Revenue Stream*** Selling refined information (HD map) for autonomous driving, location based services ***Cost Structure*** Personnel, data transfer (automotive industry)
Data offering	Sensor data (rainfall, road surface texture)	Temperature, GPS (latitude, longitude)	RSD (camera data: edge markings, center markings, strip width, crash barriers, guide posts, signs, wild warning reflectors and barriers)
Possible Use Case	Local Hazard Service	Solar Energy Prognosis	HD-Map

In the business structure analysis, data is divided into four categories. "Internal rational data" is owned by the organization and updated at certain points of time. "Internal continuous data" is owned by the organization and available in the form of continuous streams. "External rational data" is owned by third parties and updated at certain points of time. "External continuous data" is owned by third parties and available as continuous streams. This classification is used for the data contribution factor defined in Section IV.D.

B. Monetary Effect

Next, the financial perspective is taken and the monetary effect of the three use cases is considered.

Weather service

Table III shows a forecast for the sales figures for fully and partially automated vehicles [7].

TABLE III. SALES FORECAST FOR (PARTIALLY) AUTONOMOUS VEHICLES

Year	**2014**	**2020**	**2035**
Sales figures in millions	0.8	3.3	28

With an assumed willingness to pay for the Hazard Warning Service of 10 € per (partially) autonomous vehicle per year and the estimated globally available 3.3 million high or fully autonomous vehicles in the year 2020, this results in a potential revenue of 4.95 million €. The calculation assumes a market share of 15%:

$$3.3\ Mio * 10€ * 0.15 = 4.95\ Mio\ €$$

Market shares of up to 25% are often predicted in this product category [36]. In this best-case scenario, the weather service provider reaches a market share of 3.3 Mio * 0.25. This number is based on market data for navigation charts [36]. In this case, the data buyer reaches a revenue of 8.25 million €.

$$3.3\ Mio * 10€ * 0.25 = 8.25\ Mio\ €$$

For the further demonstration of the methodology the lower potential revenue of 4.95 million € is assumed.

Transmission system operator

The data sets offered in this use case are generated by a dynamic car-sharing fleet of 650 vehicles, covering the entire business area of Munich (82 km^2). The value of the data is analyzed on the one hand by means of a pure cost estimate and on the other hand by an opportunity analysis. In total, the transmission system operator allocates 1 billion € to grid stabilization in Germany in 2017 [18][34]. According to Statista, the total turnover of the German network operators in 2015 was 31.2 billion €, in which the company in question realized 3.3 billion [24][34]. This results in a market share of 10.6%.

$$\frac{3.3\ Bio\ €}{31.2\ Bio\ €} = 10.6\%$$

In 2017, 547 TWh of electricity were produced in Germany [9]. Adjusted for the market share of 10.6%, 57.8 TWh of electricity are attributable to the company.

$$547\ TWh * 10.6\% = 57.8\ TWh$$

Since the company also has another location, the total transfer performance must be considered. At the second location, the company is the sole network operator [34]. There are about 106 TWh of electricity generated [11]. This results in a total transmission capacity of 153.8 TWh per year.

$$57.8\ TWh + 106\ TWh = 153.8\ TWh$$

The necessary data for the solar energy forecast (temperature and GPS coordinates) resulting from business model analysis are delivered via the dynamic vehicle fleet in the Munich area. The power consumption in Munich is identified as a reference. In Munich, 2.8 TWh of electricity are consumed per year in total of all inhabitants [30][37]. Compared to the total transfer performance of the company, this results in a share of 1.7% for the Munich area.

$$\frac{2.8\ TWh}{153.8\ TWh} = 1.7\%$$

If one multiplies this percentage with the amount of one billion €, which the network operator identifies as costs for network stabilization in 2017 [34], this results in a maximum possible cost reduction of 17 million euros, which the network operator saves in a best-case scenario, in which perfect network stabilization over the data supply is provided.

$$1\ Bio\ € * 1.7\% = 17\ Mio\ €$$

However, it can be assumed that the provided data cannot be expected to improve the forecast of renewable energy performance for grid stabilization perfectly. Nevertheless, even the assumption of a ten percent improvement leads to a cost reduction of 1.7 million €.

This achievable value is calculated taking into account the average solar energy generated in Bavaria. The value of the data material is justified by consideration of an opportunity value. Here, the value of data for a perfect solar energy forecast results from the price and the amount of solar energy in the Munich area, which would have to be bought in the absence of such a data-based demand forecast. In other words, the amount "price times quantity" reflects the short-term purchase of solar energy which is necessary for grid stabilization, resulting from inadequate predictions. This could be avoided by making good use of the data. The Fraunhofer Institute gives an average value of produced solar energy of 9.8% in electricity generation in Germany [13]. The Federal Network Agency's statistics show that Bavaria produces the most solar energy in Germany [8]. Munich is the leader in hours of sunshine [19], which is why for Munich the average value of produced solar energy is calculated as the lower limit in a minimum scenario.

$$2.8\ Twh * 9.8\% = 0.27\ TWh$$

Taking into account a green electricity price of 56 € per MWh, this results in a potential saving of 14.4 million € (assuming a perfect prediction) [10].

It is striking that the calculated values of the best possible savings are relatively close (14.4 million € and 17 million €).

Navigation maps provider

The calculation for this use case is similar to the monetary revenue estimate for the weather service. The willingness to pay for the HD card in the automotive industry is 60 € for a highly autonomous or fully autonomous vehicle per year.

For the year 2020, a global volume of 3.3 million high or fully autonomous vehicles is expected on the market, in 2035 a total fleet of 28 million vehicles (see Table III).

The market for navigation charts in vehicles is divided into market shares of 15% to 25% [36].

These assumptions lead, in the worst case scenario (assuming a market share of 15%), to a potential turnover of at least 29.7 million € in 2020 (3.3 million * 60 €* 0.15) and 252 million € in 2035 (28 million * 60 € * 0.15).

C. Market competition

The market position is qualitatively described for all three use cases by the criteria according to [23] "is it valuable", "is it rare", "is it hard to imitate" and "is the firm organized for success". The answers to the questions are based on a competitive analysis in which 22 competitors are considered in the case of the transmission system operator, 24 for the weather service and 26 for the navigation maps supplier. The qualitative assessment shows that there is a "short term competitive advantage" for all three use cases. In the case of the transmission system operator, also a quantitative value can be set by an "in-house manufacturing vs. external purchase analysis", which means a counter-calculation of an internal generation of data as an alternative to an external purchase. From practical experience, a rigid sensor system of 2000 sensors for the price of 145 € per sensor is assumed. Surcharges are 8,000 € for housing development, 90,000 € for programming and 5,000 € for position planning.

This results in investment costs of 393,000 euros:

$$2000 * 145\,€ + 8{,}000\,€ + 90{,}000\,€ + 5{,}000\,€ = 393{,}000\,€$$

Furthermore, a monthly failure rate of 0.5% is assumed. This results from damage done by weather and vandalism to publicly accessible sensors. Corresponding monthly repair costs are 1,850 €. For data transmission 20 € per month are estimated. This results in a total annual expenditure of 415,440 €:

$$393{,}000\,€ + 12 * 1{,}850\,€ + 12 * 20\,€ = 415{,}440\,€$$

This consideration is a minimum estimate. It can be assumed that the transmission system operator does not own 2,000 relevant properties in Munich to set up sensors.

The value of 415,440 € corresponds to the amount that the transmission system operator would have to raise to collect perfect data in the same quantity.

D. Data quality and contribution factors

In order to evaluate the data quality, 14 data scientists are interviewed in four expert workshops for each case on given quality criteria. The results of the workshops are summarized in Figure 2 and transformed into a quality factor Q.

$$Q = \sum_{i=1}^{n} c_i * w_i \qquad (10)$$

Q Quality factor
c_i Evaluation factor of criterion i
w_i Weight of criterion i
i Criterion index
n Number of criteria

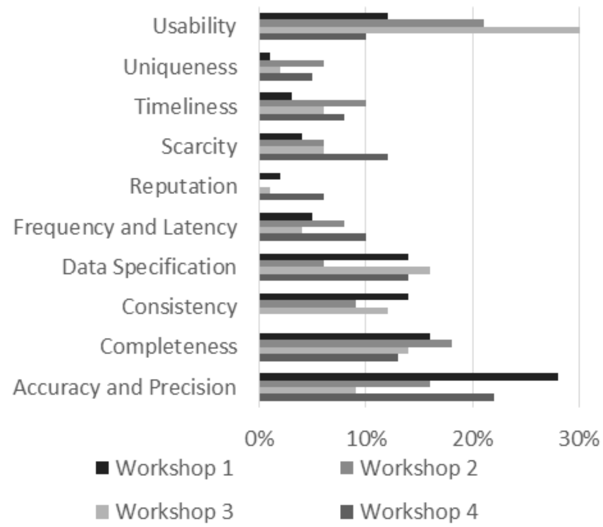

Figure 2. Evaluation of the quality criteria of data

The evaluation factor values c_i of the criteria result from a pairwise comparison of all criteria in a preference matrix. The data contribution factor *C* expresses if all required data (contribution factor = 1), almost all data (contribution factor = 0.75), about half of the data (contribution factor = 0.5), few data (contribution factor = 0.25) or no data at all (contribution factor = 0) can be provided.

$$C = CM * w_{cm} + CK * w_{ck} \qquad (11)$$

C Data contributing factor
CM Contribution factor of metadata
CK Contribution factor of key data
w_{cm} Evaluation of metadata
w_{ck} Evaluation of key data

The use-case-specific factors CM und CK differentiate between key data und metadata (additional data). The expert based weights w_{cm} and w_{ck} rate the relative data contribution of each data type in the use case. Location based services are one example. They are based on customer preferences (metadata) on the one hand and GPS data (key data) generated by the vehicle on the other hand. Experts give weights of 0.2 for GPS data and 0.8 for preference data. It is assumed that the vehicle can deliver 90 percent of the GPS data needed, however none of the preference data. So, the contribution factor of the metadata is zero. The calculation of C results in the value of 0.18:

$$0.2 * 0.9 + 0.8 * 0 = 0.18$$

The corresponding calculation rules provide the following results (see Table IV) coming from the workshops for the three use cases.

TABLE IV. WORKSHOP RESULTS

Key figure	Quality factor	Contribution factor
Weather service	0.79	0.5
Transmission system operator	0.8	1.0
Navigation maps provider	0.8	0.55

E. Selling prices

Through the product of monetary effect (*V*), quality (*Q*) and contribution factor (*C*) (see Figure 1) a value for the offered data of 1.95 million € (4.95 million € * 0.79 * 0.5) for the use case “weather service” is calculated, for the use case “transmission system operator” 332,352 € (415,440 € * 0.8 * 1.0) and for the use case "navigation maps manufacturer" in the worst case scenario 13.07 million € (29.7 million € * 0.8 * 0.55). With the weather service as well as the navigation maps manufacturer the turnover is in the foreground. In the case of the transmission system operator, the focus is not on turnover, but on cost savings. The costs for the self-collection of the data are less than the calculated costs of externally purchased data that lead to the same savings effect, so that this lower value of about 415,440 € is used. In the case of self-collection of data the contribution factor is 1.0, since all required data is recorded internally.

F. Lessons Learned

On the one hand the presented evaluation model was applied to several ficitious use cases with real information coming from companies interested in buying data but without any decision. On the other hand, the evaluation model was tested on several specific sales situations and the outcome of the model, i.e., the post calculated data value, was compared to the real sales price. The monetary data values determined using the evaluation model show an average deviation of 8% from the sales prices negotiated in practice.

V. CONCLUSIONS

The methods of data evaluation identified in the literature are individually not suitable for practical value determination of data and their pricing in sales situations. This paper presents a methodology that focuses on the selling of data as intangible products to external business partners.

The methodology can also be transferred to use cases within the company. In addition to determining the value of the data, decisions regarding the pricing model must be made.

However, for long-term strategies it is unclear to what extent recorded data is valuable in the future. Data that is still useless, because currently there are no use cases, can be highly relevant for future use cases. Because of the existing knowledge gap and missing empirical values, it is impossible to determine a value of data over the entire lifecycle, above all because of very uncertain future potentials.

This article exemplifies a possible evaluation and monetization of a small fraction of the total data available in the automotive industry.

Looking at the huge amounts of data available there, it quickly becomes clear that due to technical limitations probably never all potential use cases can be implemented. There are various transmission options for vehicle generated data. The built-in memory can be read in authorized garages, updates can be transmitted at weekly or daily intervals, or data can be transferred in real time. There is always a technical limitation due to the restricted transfer rates or transfer options. Not all conceivable applications can be realized at the same time.

It is also an open question whether it makes sense to regard the vehicle as an open platform. In this case, an automobile manufacturer or even the automotive industry as a platform provider could probably sell the platform as a service (PaaS) to service providers who will pay for specific data accessed via the platform. As an analogy, platforms of Apple and Android can be considered. Third parties develop services to be offered on these platforms. The developed services (e. g., apps) increase the attractiveness of the platform. Depending on the design, there are direct and indirect network effects. With regard to autonomous driving, this approach may potentially increase the attractiveness of vehicles and vehicle fleets acting as such platforms. For example, at BMW, in addition to many existing Connected Drive services, applications of third-party providers can be activated, which leads to an immense increase of the value of a ride and the driving experience for the customer. Here, completely new service ecosystems spanning and connecting different industrial sectors are appearing. To name only one step toward the future, the intelligent personal assistant from BOSCH enables the networking of car services and e-home services [6].

REFERENCES

[1] F. Bodendorf, T. Meissner and J. Franke, “Valuating and Pricing of Vehicle Generated Data as a Marketable Product in the Automotive Industry,” The Eighth International Conference on Advances in Vehicular Systems, Technologies and Applications (VEHICULAR 2019), Jun. 2019, pp. 16-21, ISSN: 2327-2058, ISBN: 978-1-61208-720-7

[2] J. Balasubramanian et al., “Car data: paving the way to value creating mobility : Perspectives on a new automotive business model,” Advanced Industries McKinsey, 2016

[3] L. A. Bettencourt and A. W. Ulwick, “The customer-centered innovation map," Harvard Business Review, vol. 86, pp. 109-114, 2008

[4] BMW. *BMW ConnectedDrive Customer portal - digital networking to your BMW*. [Online]. Available from: https://www.bmw-connecteddrive.de/app/index.html#/portal/store/Base_TolCarOffer. Retrieved [05-31-2020].

[5] BMW. *NOW mobility services* . [Online]. Available from: https://www.bmwgroup.com/de/marken/now-mobilitaetsdienstleistungen.html. Retrieved [05-31-2020].

[6] BOSCH. *Only driving was yesterday - the personal assistant is tomorrow*. [Online]. Available from: https://www.bosch-presse.de/pressportal/de/de/nur-fahren-war-gestern-%E2%80%93-der-persoenliche-assistent-ist-morgen-101568.html. Retrieved [05-31-2020].

[7] A. Brugger. *Global production forecast for semi- and fully automated vehicles*. [Online]. Available from: https://www.oliverwyman.com/con-tent/dam/oliver-wyman/global/en/files/who-we-are/press-releases/Oli-verWyman_Graphics_Value%20Pools%20Autonomous%20Driving_EN_16072015_final.pdf. Retrieved [05-31-2020].

[8] Bundesnetzagentur. *Figures, data and information on the EEG*. [Online]. Available from: https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/ErneuerbareEnergien/ZahlenDatenInformationen/zahlenunddaten-node.html. Retrieved [05-31-2020].

[9] B. Burger. *Electricity generation in Germany in 2017*. [Online]. Available from: https://www.ise.fraunhofer.de/content/dam/ise/en/documents/News/Stromerzeugung_2018_2_en.pdf. Retrieved [05-31-2020].

[10] e-control. *Current market price pursuant to § 41 Green Electricity Act*. [Online]. Available from: https://www.e-control.at/en/industrie/oekoenergie/oekostrommarkt/marktpreise-gem-paragraph-20. Retrieved [05-31-2020].

[11] L. Eglitis. *Energy balance in the Netherlands*. [Online]. Available from: https://www.ebn.nl/en/the-energy-balance-in-the-netherlands/. Retrieved [05-31-2020].

[12] J. L. L. Francisco and J. Esteves, “Value in a Digital World: How to assess business models and measure value in a digital world,” Springer, 2017, ISBN 978-3-319-51750-6

[13] Fraunhofer. *Share of renewable energies | Energy Charts*. [Online]. Available from: https://www.energy-charts.de/ren_share_de.htm?year=2018&source=solar-share&period=monthly. Retrieved [05-31-2020].

[14] W. Gründinger. *Data-Driven Business Models in Connected Cars, Mobility Services & Beyond.* [Online]. Available from: https://www.bvdw.org/fileadmin/bvdw/upload/publikationen/connected_mobility/20180418_data_driven_business_models_Seiberth_Gruendinger.pdf. Retrieved [05-31-2020].

[15] R. Harmon, H. Demirkan, B. Hefley and N. Auseklis, “Pricing Strategies for Information Technology Services: A Value-Based Approach” 42nd annual Hawai'i International Conference on System Sciences (HICSS 2009), IEEE Press, Jan. 2009, pp. 1–10, ISBN: 978-1-424-44197-6

[16] N. Henke et al., “The Age Of Analytics: Competing In A Data-Driven World,” McKinsey Global Institute, 2016

[17] HERE. *Location-Based Services for the Autonomous Future.* [Online]. Available from: https://www.here.com/en. Retrieved [05-31-2020].

[18] M. Holland. *Power grid under pressure: Tennet reports record costs for emergency interventions*. [Online]. Available from: https://www.heise.de/newsticker/meldung/Stromnetz-unter-Druck-Tennet-meldet-Rekordkosten-fuer-Noteingriffe-3929093.html. Retrieved [05-31-2020].

[19] M. Kniekamp, Spitzenreiter. *In this statistic Munich is Germany's showcase city*. [Online]. Available from: https://www.merkur.de/bayern/muenchen-ist-laut-studie-zdf-deut-sche-stadt-mit-meisten-sonnenstunden-9855027.html. Retrieved [05-31-2020].

[20] D. B. Laney, “Infonomics: How to monetize, manage, and measure information as an asset for competitive advantage,” Routledge, 2018, ISBN: 978-1138-09038-5

[21] E. Leonidas, DataStream. *A Practical Guide to Pricing Data Products.* [Online]. Available from: https://cdn2.hubspot.net/hubfs/573334/Downloadable_Content_(WP_or_Guides)/DataStreamX_Data_Product_Pricing_Whitepaper.pdf. Retrieved [05-31-2020].

[22] C. Lim et al., “From data to value: A nine-factor framework for data-based value creation in infor-mation-intensive services,” International Journal of Information Management, vol. 39, pp. 121–135, 2018, ISSN: 0268-4012

[23] S.M. Liozu, W. Ulaga, “Monetizing Data, A Practical Roadmap for Framing, Pricing & Selling Your B2B Digital Offers,” VIA Publishing, 2018, ISBN: 978-1-945815-04-1

[24] D. Loesche. *Infografik: Comparison of the four German transmission system operators*. [Online]. Available from: https://de.statista.com/infografik/6023/vergleich-der-vier-deutschen-eebertragungsnetzbetreiber/. Retrieved [05-31-2020].

[25] K. Mathis and F. Köhler, “Data-Need Fit - Towards data-driven business model innovation” The Fifth Service Design and Innovation conference (ServDes.), pp. 458–467, 2016, ISBN: 978-91-7685-738-0

[26] D. Mcglivary, “Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information: Definitions of Data Categories” Chief Data Officer & Information Quality Symposium, 2009

[27] T. Meißner, “Value and utilization of data on hardware and software usage in the vehicle,” University of Erlangen-Nuremberg (Master Thesis), 2018

[28] K. O'Neal, “Quantifying the Value of Data: The First Step in Data,” unpublished presentation

[29] A. Osterwalder and Y. Pigneur, “Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers, Hoboken,” John Wiley & Sons, 2010, ISBN: 978-0470-87641-1

[30] Photovoltaik Rechner. *Information, Calculation & Companies*. [Online]. Available from: https://www.rechnerphotovoltaik.de/photovoltaik/in/bayern/muenchen. Retrieved [05-31-2020].

[31] B. Schmarzo and M. Sidaoui. *Applying Economic Concepts To Big Data To Determine The Financial Value Of The Organization’s Data And Analytics, And Understanding The Ramifications On The Organizations’ Financial State-ments And IT Operations And Business Strategies*. [Online]. Available from: https://infocus.delltechnologies.com/wp-content/uploads/2017/04/USF_The_Economics_of_Data_and_Analytics-Final3.pdf. Retrieved [05-31-2020].

[32] C. Shapiro and H. R. Varian, Versioning, “The Smart Way to Sell Information,” Harvard Business Review, vol. 6, pp. 107 f., 1998

[33] F. Stahl and G. Vossen, “High quality information provisioning and data pricing” 29th International Conference on Data Engineering Workshops (ICDEW 2013), 2013, pp. 290-293

[34] TenneT Holding B.V.. *Annual report 2017 TenneT Holding B.V.* [Online]. Available from: https://annualreport.tennet.eu/2017/annualreport. Retrieved [05-31-2020].

[35] J. H. Thun, “Maintaining preventive maintenance and maintenance prevention: analysing the dynamic implications of Total Productive Maintenance,” System Dynamics Review: The Journal of the System Dynamics Society, vol. 22, pp. 163-179, 2006

[36] S. Viswanathan and G. Anandalingam, “Pricing strategies for information goods,” Sadhana Indian Academic Sciences Journal, vol. 30, pp. 257–274, 2005, ISSN: 0256-2499

[37] Wirtschaftswoche. *GPS-Satellitennavigation: Wie Navi-gationsgeräte-Firmen um einen Milliarden-Markt kämpfen.* [Online]. Available from: https://www.wiwo.de/unternehmen/gps-satellitennavigation-wie-navigationsgeraete-firmen-um-einen-milliarden-markt-kaempfen-seite-3/5426542-3.html. Retrieved [05-31-2020].