

Scientific Approaches and Methodology to Determine the Value of Data as an Asset and Use Case in the Automotive Industry

Frank Bodendorf
Friedrich-Alexander-University of
Erlangen-Nuremberg (FAU)
frank.bodendorf@faps.fau.de

Klaus Dehmel
Friedrich-Alexander-University
of Erlangen-Nuremberg (FAU)
klaus.dehmel@fau.de

Jörg Franke
Friedrich-Alexander-University of
Erlangen-Nuremberg (FAU)
joerg.franke@faps.fau.de

Abstract

From a theoretical perspective data does not constitute a traditional business asset. Existing valuation approaches are either sector specific or still unexplored. In modern businesses the value-adding use and monetization of existing "big data" represents one of the greatest potentials in the context of digital transformation. This paper aims at reviewing methods and developing an integrated methodology for the value determination of data in general and for use in the manufacturing industry in particular. Therefore, the general state of research in data value assessment is investigated by a broad literature analysis. Based on the identified general principles, methodological requirements for data value determination are compiled. A new methodology for data evaluation is developed and applied to four use cases coming from the automotive industry. The results show that the methodology can be used in different contexts and thus enables managers to explore the most promising use cases for data-driven business.

1. Introduction

For many companies owning data is already more important than holding traditional assets [1]. From a traditional asset theory perspective, however, data does not constitute an asset. As a result, data and its value in business practice are generally not structurally treated and valued as such [2]. In contrast, the accounting and maintenance of traditional financial assets such as fixed assets, real estate, etc. is standardized and established in organizations [3]. According to the quote "if you can't measure it, you can't manage it", however, the valuation of resources represents the most important step for their goal-oriented management and control [4]. Since the end of the 1990s there has also been a direct need in the field of data management to quantify the value of data for an organization analogously to other tangible and intangible resources [5]. Although the demand for monetary data valuation has grown continuously over the course of the worldwide increase in data volumes,

no generally applicable, cross-sector methods have yet been established for this in theory and practice [6]. Existing calculation approaches are either industry specific solutions or still largely unexplored in practice [3]. The problem is also illustrated by a worldwide survey conducted by Gartner Consulting in 2014: According to this study, 89% of the organizations already regard data as structural assets, but only 24% can quantify their actual financial value [7]. Since traditional business management approaches to controlling assets can be applied to data only to a very limited extent, this methodological gap often leads in practice to the use of methods that are not suitable for data evaluation [8].

The automotive industry is particularly affected by the disruptive innovation effects of data valuation [9]. The value-adding use of large amounts of data holds one of the greatest transformation potentials for the industry. A modern car generates up to 25 gigabytes of data per hour [10]. In combination with innovative analytics tools these data volumes hold a competitive potential for manufacturers both for the development of customized products and new business models as well as for process optimization. E. g., a McKinsey study estimates the added value of profitable use and monetization of data generated in vehicles at 450 billion US dollars by 2030 [11].

This paper makes a contribution to those studies from a scientific perspective. Section 2 takes a short look at the theoretical background of data valuation. Section 3 examines the practical suitability and up-to-dateness of methods mentioned in literature. Section 4 introduces approaches to data valuation to prioritize data-driven business use cases. On this basis we propose a new methodology for holistic data valuation and priority setting which is adapted to the requirements of today's data business. Finally, sections 5 and 6 validate and conclude the research by looking at real world use cases coming from a large automobile manufacturer.

2. Data Valuation Review

2.1. Determining factors of data value

Data can generate value for an organization in different ways. On the one hand, data can generate a specific value for an organization with regard to its specific business model. On the other hand, data can also be sold as a marketable product. Gartner strategy consultants use this classification to distinguish between direct and indirect data monetization [12]. Value creation based on data can also be considered a four-phase maturity model:

- o Monitoring (Phase 1): Data is used to detect potentials to improve the efficiency of existing processes and simultaneously identify their costs and risks.
- o Optimization (Phase 2): Data is used to gain a better understanding of the customer in order to optimize products and services.
- o Growth (Phase 3): Data is used to generate new sales opportunities and new sources of revenue for an organization.
- o Monetization (Phase 4): In this phase, data itself is priced as a primary product on the market (highest level of value creation).

The true data value only arises looking at the entirety of all phases of the data value chain. Thus, the factors determining the internal value of data are equally relevant as the parameters determining the value on data marketplaces. A research report by the Fraunhofer Institute shows that two main parameters influence the value of data [13]: context and intended use. The first factor considers the circumstances and conditions under which the data is used. This is further detailed by the parameters data source, quality, and type. The factor "purpose of use" includes the market on which the data is traded as well as the actors who provide or require the data. Further influencing factors are law-related (context) and economy-related (use) ones [14]. Law-related factors depend on aspects of data usage or data transfer. Economic factors include the market form or the exclusivity of data access [15].

According to a study conducted by the University of Potsdam in 2013, data quality is the most important influencing factor for the successful value creation through data-driven analytics and thus for the value of data [16]. This finding is supported by the results of a study done by the IT consultancy MHP, which states that securing and improving data quality are the most important tasks in value-oriented data management. It is not without reason that veracity – the quality of data – is the sixth dimension of the definition of “Big Data” along with volume, variety, velocity, variability, and value itself [17]. Veracity comprises two aspects: first, the

consistency of the respective data sample which means their statistical reliability. Second, the aspect of data trustworthiness which is influenced by several factors like the origin of data and its collection and processing methods. The definition of veracity shows that the quality factor is one of the most complex parameters in data value determination, which is composed of a large number of other parameters [18]. Therefore, data quality is discussed in more detail below. Neither in the literature nor in business practice there is a comprehensive list of factors influencing data quality. Since the term is strongly subjectively coined, many definitions remain on a very simplified level, e. g., saying that data shows a high level of quality if it is suitable for its intended purpose in enterprise, decision-making, and planning [19]. Data quality is also influenced by corresponding factors (also called quality attributes or dimensions) depending on the specific application [20]. In the scientific literature you find a broad spectrum of these factors of data quality. Already in 1995, Wang et al. present a total of 15 data quality dimensions, which are extended in 2001 by Redman to a total of 51 quality attributes [18] [21]. Figure 1 illustrates the procedure of quality criteria assessment.

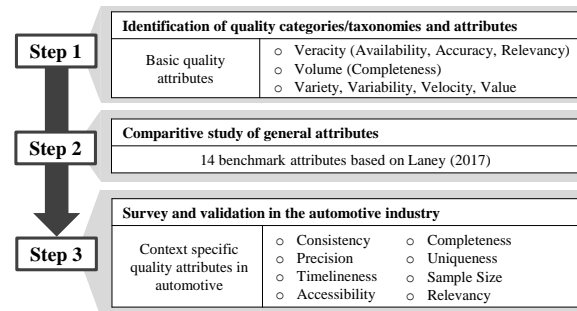


Figure 1. Exploration of data quality criteria.

In the first step, general quality categories are assigned via classification schemes, so-called taxonomies. A literature review of the most important, generally applicable taxonomies for data quality shows that the attributes "Accuracy", "Completeness", "Relevancy", and "Availability" are largely included across categories [22]. In the second step, the previously identified factors are compared to quality attributes relevant in the context of data valuation. A total of 14 quality attributes mentioned in the Gartner concept of "Infonomics" are used as a basis [12]. In the final step, these parameters are compared to quality attributes specifically used in the automotive industry and validated for their practical applicability. This is done by a survey that is conducted among data specialists of a German premium car manufacturer [23]. The quality attributes found in this three-steps approach are considered relevant in the context of this paper.

Data quality within the framework of data evaluation is a multidimensional concept. For an organization, this concept consists of both subjective perceptions of individual persons and objective, process-specific criteria [24]. Subjective quality criteria on the one hand can only be defined by individual users based on their personal views, experiences, and backgrounds. Objective criteria, on the other hand, can be determined precisely on the basis of a detailed analysis of the data.

2.2. Scientific approaches to determine data value

After looking at general principles and factors influencing the value of data in Section 2.1., this section focuses on methods and models for specific valuation approaches that can be extracted from scientific literature. Three financial valuation approaches are associated with market value, cost, and benefits [25]. In addition, recent publications recommend a risk-oriented approach to data valuation because of the large monetary business risks that can arise for an organization when data is lost or misused [26]. In cost-oriented valuation approaches, the costs that the data owner must bear for production, procurement, and maintenance form the data value. Benefit-oriented approaches calculate a data value based on the financial returns arising from data use over the entire life cycle. The financial value of data is reflected by the contribution data can make to business performance, such as increasing customer satisfaction. However, in benefit-oriented approaches, data generates value for an organization not only financially, but also from a qualitative perspective. In the reviewed literature there is a consensus that the utility value of data should also consider non-financial qualitative perspectives [27]. In risk-oriented approaches, the data value is derived from the monetary damage potential regarding the data quality required for the application and the effort required to improve and manage data quality errors [28]. Market value oriented approaches are based on the assumption that data is sold on data markets and that sales prices ultimately determine its value. The different approaches are structured in Figure 2 where transparent icons characterize use case specific approaches.

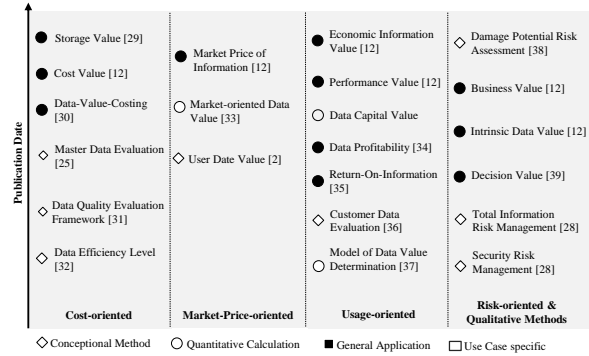


Figure 2. Data valuation approaches in scientific literature.

3. Evaluation Methods Review

In this section, the approaches shown in Figure 2 are evaluated and analyzed regarding their suitability for a holistic data value determination in the automotive industry. For this purpose, corresponding specific requirements for data value determination are defined. This procedure follows the standard process of requirements engineering defined by the IEEE association, which consists of four steps, shown in Figure 3 [40].

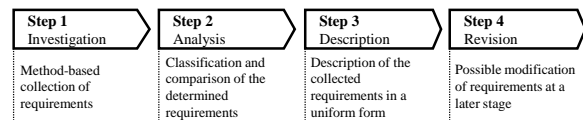


Figure 3. Requirements engineering process.

Step 1 is based on the inventory approach. This is one of the most well-known methods of requirements analysis in the software area and done by studying written documents such as reports or statistics. For this step a literature analysis is conducted here. In step 2 of the process the requirements identified in step 1 are categorized in two different ways. On the one hand, they are differentiated according functional and non-functional requirements [41]. On the other hand, these functional and non-functional requirements are assigned to three main requirement groups A, B, and C.

A) Holistic approach of data value

One of the main factors influencing data value is the data quality, which can be measured by different criteria [42]. The costs incurred by data storage, preparation, and maintenance over the entire data value creation cycle reduce the data value. For a holistic data evaluation, a calculation logic must include cost factors in addition to benefit factors. In total, four functional requirements belong to Group A:

- o Requirement A1: The calculation logic of the methodology takes into account a variety of different

financial and non-financial parameters that influence data value.

- o Requirement A2: The calculation methodology does not only consider monetary benefits of data, but also includes calculation factors that measure the value of data in a qualitative way.
- o Requirement A3: All data quality factors identified in Section 2.1 must be included in the methodology.
- o Requirement A4: Cost parameters incurring over the whole data life cycle have to be considered in the methodology.

B) Laws of data value

The literature review reveals seven laws that describe differences between the value theory of data and the value theory of material goods. The principles of divisibility of data and time-dependent loss of data are taken into account. Moreover, data does not have a value that is realized only once, but can generate its value over a value creation lifecycle. In Group B, this lifecycle aspect results in the following two functional and one non-functional requirements:

- o Requirement B1 (functional): Integration of the simultaneous value creation potential of data for different users into the evaluation logic
- o Requirement B2 (functional): Integration of factors that reflect the time dependency of the data value into the evaluation logic
- o Requirement B3 (non-functional): Integration of the overall value creation potential into the evaluation logic of the methodology, not only considering just the data value within a use case

C) Prioritization of data value

In order to identify and prioritize the most relevant use cases of data as an asset, the methodology must be based predominantly on calculation factors that can be determined with little effort. While the expected benefit as well as costs usually require time consuming estimation procedures, the data quality of an existing data basis, for example, can be determined at an early stage.

Prioritizing data-driven use cases allows also to make general statements about the overall data potentials. One approach is to consider the value contributions of use cases realized in the past as the basis for future data value (see approaches described in Section 2.2). However, this approach has a fundamental structural disadvantage. Potentials realized in the past are often achieved in an environment and with a data infrastructure that is not primarily designed for data-driven use cases. However, to justify investments in strategic data collection, storage, and maintenance, the data infrastructure available for future vehicles must be

considered. This leads to the following requirements within Group C:

- o Requirement C1 (functional): Consideration of parameters that can be determined and quantified at an early stage
- o Requirement C2 (non-functional): Avoidance of data prioritization based on past use cases and former data infrastructures

In expert workshops the evaluation methods identified in the literature are reviewed using a scoring approach to assess the degree of fulfilment of the mentioned requirements. In the analysis all approaches described in Section 2.2 are considered, which actually determine the data value in a quantitative and general way (visualized by black bullets in Figure 2). Figure 4 reflects the extent to which these methods meet the requirements as a percentage value. In this way, Figure 4 shows which of the three basic approaches (cost-oriented, benefit-oriented, and further methods) meet the requirements of the three individual groups best, when viewed as a whole.

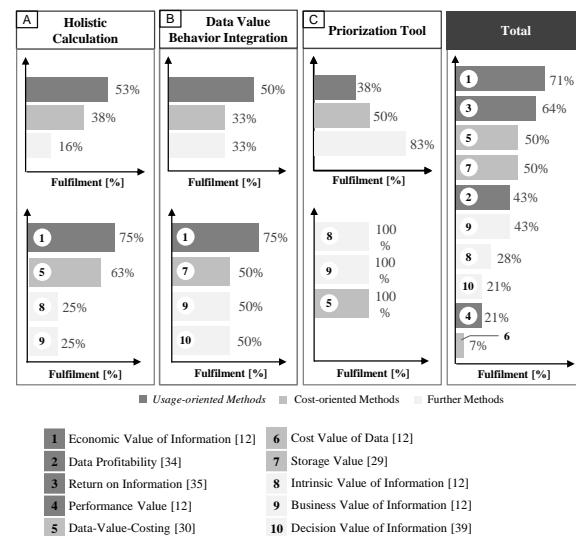


Figure 4. Fulfilment of data evaluation requirements

Finally, the graph shows how the methods perform across all requirements (right column). Usage-oriented methods are best suited for a holistic data evaluation approach with a degree of fulfilment of 53 % (requirements of Group A). With a degree of fulfilment of 50 %, they also take into account the special characteristics in data value behavior (requirements of Group B). To prioritize data-driven use cases (requirements of Group C), it is recommended to use the instruments included in the category "further methods" (degree of fulfilment 83%). In addition, the analysis shows that none of the methods fulfills all requirements sufficiently at the same time. None of the methods

considered is fully suitable (100%) for a holistic data evaluation. Laney's economic value approach achieves the maximum percentage (71%), but is far away from the optimum.

4. Methodology for Data Valuation

4.1. Data value analysis approach

To integrate the four-stage data value creation logic (see Section 2.1) into the methodology to be developed, the concept of maturity assessment for car components, established in the automotive industry, is used. It measures the maturity of components with regard to their suitability for series production at defined milestones in the product development process.

This logic is also applicable in the context of data quality determination. First, the analysis framework is comparable. While measuring the maturity of material components for manufacturability in series production, it is crucial for data quality determination to find out to what extent the required data has the appropriate maturity needed for the manufacturing use case. Second, the value-adding use of data follows the value creation process in manufacturing. In this process, data must meet more stringent quality criteria as the value-added process progresses. In addition, a maturity assessment is generally advantageous if the associated results cannot be made measurable by means of key figures. Finally, the logic of data maturity assessment is already established in the automotive industry, which elevates the acceptance of the proposed methodology in practice.

While transferring the logic of maturity assessment to the area of data quality, a first step is to define relevant maturity levels g_M (g: degree; M: maturity level).

Maturity level 1: Technical validity of data (g_{M1}).

The first maturity level indicates whether the data examined for a use case meets the purely technical requirements for data evaluation and is therefore "technically" valid for this use case. This is expressed by the question whether the systems are connected (connected systems $g_{M1_{Co}}$), whether the data has an appropriate format ($g_{M1_{F}}$) and whether the data is accessible to all parties involved (accessibility $g_{M1_{Ac}}$).

$$g_{M1} = \sqrt[3]{g_{M1_{Co}} \times g_{M1_{F}} \times g_{M1_{Ac}}} \quad (1)$$

Maturity level 2: Business relevance of data (g_{M2}).

This level of maturity represents the relevance of the examined data for the intended purpose. On this level, it is checked whether the data can be used (usability $g_{M2_{P}}$) and whether the data is available in sufficient sample size (sample size $g_{M2_{S}}$).

$$g_{M2} = \sqrt[2]{g_{M2_{P}} \times g_{M2_{S}}} \quad (2)$$

Maturity level 3: Business readiness of data (g_{M3}).

The third level of maturity reflects the extent to which the data is capable of generating value within the application framework of a use case. The parameters relevant for calculation are the consistency of the data ($g_{M3_{CT}}$), its accuracy ($g_{M3_{A}}$), timeliness ($g_{M3_{T}}$), and completeness ($g_{M3_{C}}$). The individual quality factors are rated by the responsible data scientist using a four-level scale (1: poor proficiency to 4: very high proficiency).

$$g_{M3} = \sqrt[4]{g_{M3_{CT}} \times g_{M3_{A}} \times g_{M3_{T}} \times g_{M3_{C}}} \quad (3)$$

The total data maturity level $g_{M_{total}}$ is aggregated from the individual maturity levels using the same calculation.

$$g_{M_{total}} = \sqrt[3]{g_{M1} \times g_{M2} \times g_{M3}} \quad (4)$$

4.2. Cost determination approach

A detailed cost calculation in each of the four phases (see Section 2.1) is not possible. Use case specific costs can only be identified for phase 4. However, according to internal calculations in company specific case studies the costs incurring in the fourth phase represent 70 % to 95 % of the total costs over the entire data management life cycle.

Thus, the first step is to find a suitable approach that can be used to determine the costs for Phase 4. It is assumed that the management of data analysis projects is comparable to the management of complex software or IT projects [43]. Here, experience-based cost estimation by expert interviews, algorithmic cost modelling, analogical reasoning, Parkinson's Law or pricing-to-win are popular methods established in practice [41]. For cost estimation of data-driven use cases in the automotive context, the method of "analogical reasoning" is used for the reasons explained below. In software projects to be reevaluated, the analogical reasoning is based on completed projects that have already been evaluated on the cost side. This makes it possible to increase efficiency in cost evaluation by simplifying the formation of analogies. In the case study here, this method fulfils all conditions, which must apply for the use of the analogical estimation. The analogical approach is based on specific comparison criteria [44]. For this purpose, the following methods of experience-based software cost estimation are analyzed: The Function Point method, the CoCoMo method, the ObjectMetrix method and the Use Case Points method, which belong to the most popular ones [45].

The Function Point method allows a quantitative estimation of software costs at an early development stage and is characterized by high adaptability and simple comprehensibility. With this method, the scope

and complexity of software are assessed and rated by function points. The prediction of the project effort is based on these function points.

The factors, which mainly influence the costs of data arising in phase 4, are determined by expert discussions. This leads to eight evaluation factors: number of data sources, data transformation effort, variance of data source type, size of data volume, number of data users, query complexity, data protection requirements, redundancy of data storage. The individual factors for a specific use case are evaluated by a data scientist responsible for the use case data. The data scientist assigns the value 3 for a high ("L"), 2 for a medium ("M") and 1 for a low ("S") expression of the factor associated with the data of the considered use case. For cost evaluation it is decisive to differentiate whether the data is stored and processed on the organization's own internal servers or on rented external servers, e. g., in a cloud. This relationship is considered via weighting the function points. Depending on this, the cost factors of "size of data volume", "data protection requirements", and "redundancy of data storage" differ (see Figure 5).

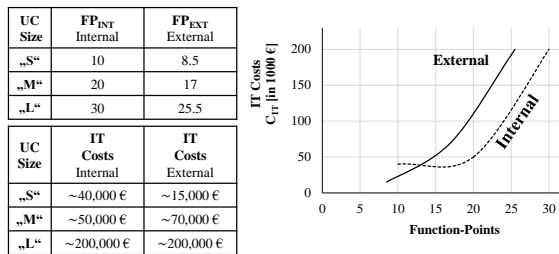


Figure 5. Cost determination of use cases (UC) by Function Point approach.

Finally, the function points of the respective valuation factor result from a line-by-line multiplication of the valuation factor X by the weighting Y. The column total of all function points characterizes the use case with regards to the cost factors. Using the before mentioned function point methodology historical data projects are evaluated in expert workshops. The interpolation of results leads to the functional relationship between function points and IT costs qualitatively shown in Figure 5 which can be described mathematically as:

$$Ext: C_{IT} (FP_{EXT}) = 0,5 \times (FP_{EXT})^2 - 6,7 \times FP_{EXT} + 35 \quad (5)$$

$$Int: C_{IT} (FP_{INT}) = 0,7 \times (FP_{INT}) - 20 \times FP_{INT} + 170 \quad (6)$$

4.3. Storage value potential approach

This methodology allows a quantitative determination of the quality of data assets. The original formula of the "data value costing method" divides the total cost incurred over the data management lifecycle by the parameters "user identification level", "timeliness", and "error rate". Since the data maturity indicator presented in the previous section is more accurate due to the large number of influencing parameters behind it and the background is very similar, the error rate parameter is replaced by the data maturity parameter introduced in the previous section. The "timeliness" parameter is included in the maturity value and therefore eliminated as a separate calculation parameter. As in the original formula, the "user identification degree g_{UI} " remains in the denominator as a factor influencing the data value. Formula 7 shows the calculation for the added storage value potential approach.

$$Added\ SVP = \left[\left(\sum_{i=1}^n g_{PI_i} \times PI_i \right) + \left(\sum_{e=1}^m g_{PI_e} \times PI_e \right) \right] \times \left(\frac{C_{total}}{g_{M_total} \times g_{UI}} \right) \quad (7)$$

The added storage value potential (SVP) of the data considers the use by further internal (formula symbol n) and external parties (formula symbol m). The plausibility factor g_{PI} is introduced for this purpose. In the case of internal parties, this describes to what extent the internal departments will benefit from the respective data (formula symbol g_{PI_i}). In the case of external parties, the parameter expresses the probability that the potential recipient of the data is actually willing to pay for this data (formula symbol g_{PI_e}). The degree of plausibility is measured in percent. The storage value potential is finally calculated by multiplying the sum of all potentials PI of internal and external parties (in each case corrected by multiplying the plausibility factor g_{PI}) by the costs C_{total} , which are modified by the factors g_{M_total} and g_{UI} . The calculated numerical value reflects the potential value of the data through multiple use in different areas of an organization. This value is derived from the costs incurred.

4.4. Future oriented data value approach

The future-oriented data value determination is done by replacing the economic data value of Laney (see Section 2) by a time-dependent function. First the data value factor is expressed in a more realistic way by introducing four data benefit categories. Data can generate financial benefits and hence value in different ways. Since the turnover parameter (symbol R) contained in the formula so far does not reflect these different benefit potentials, the parameter is renamed the benefit B. Based on the relevant financial control

variables in the automotive industry, the following benefit categories can be defined:

- o Increase of the contribution margin (formula symbol B_{CM})
- o Reduction of warranty costs (B_{WC} symbol)
- o Reduction of one-time expense costs (formula symbol B_{OTEC})
- o Reduction of manufacturing costs (B_{MC} symbol)

If data generates more sales or if data is monetized and sold as an asset, the contribution margin increases. Analogous to the logic of the economic information value, the data value results from the sum of the financial benefits realized within the different benefit categories described above. This benefit is reflected by the difference between the financial benefit resulting from the use of the data (addition $_I$ in the index) and the value without use of the data. If concrete figures are not known, the value is set to zero. As a rule, the financial benefit of the data is determined at the end of the observation period t_b in which the value is generated. The lifespan of this data usually extends well beyond the observation period t_b (except for data that quickly loses value). Therefore, the parameter B_D , which can be calculated by formula 8, describes the financial benefit of the data at the end of the observation period (time t_b).

$$B_D = (B_{CM_I} - B_{CM}) + (B_{WC_I} - B_{WC}) + (B_{OTEC_I} - B_{OTEC}) + (B_{MC_I} - B_{MC}) \quad (8)$$

The next factor included in the formula is the time dependency of the financial benefit of data. Therefore, the parameters that influence the value of data over time must be identified in the first step. The degree of aggregation of the data plays a decisive role here: Individual data loses value over the course of time, whereas data that is continuously combined with other data increases in value over the course of time. However, this does not apply equally to all data of an organization. Rather, the purpose for which the data is used also influences its value over time [24]. This dependency is investigated by a study of the American Nuclis Research Institute, among others. In a total of 47 companies, the half-lives of data are analyzed depending on whether strategic, tactical or operational data is involved [46]. Value functions take into account the time factor t for various data classes. Strategic data (VSD) has a linear value progression (see Figure 6).

The slope of the curves is negative for single data and positive for aggregated data. The value progressions of individual and aggregated tactical and operational data are similar to those of exponential functions. The following assumptions apply: Individual data reaches its maximum value at the end of the initial observation period (duration t_b). In practice, the data value in a use case is usually determined at the end of an initial

analysis period t_b . The factor T indicates the data lifetime in analogy to the economic data value. As a rule, the length of the observation period t_b and the financial data benefit $V_x(t_b)$ (for x see Figure 6) achieved therein are known.

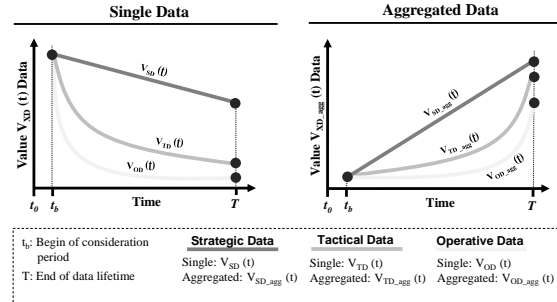


Figure 6. Time dependency of data value.

The investigations of the Nuclis Research Institute show that the value of strategic data at the end of the lifecycle is either halved (for individual data) or increased by 50% (for aggregated data) compared to the value of the data at time t_0 . In accordance with the gradient triangle of linear functions, the gradient for strategic data is shown in the formulas 10 and 11. Based on the previously made assumption that the value at time t_0 corresponds to the value at time t_b , the intersection of the gradient line $V_{SD}(t)$ with the y-axis gives the value of the data at time $V_{SD}(t_b)$ (symbol c in the general linear function, see formula 9).

$$\text{Linear Function : } y(x) = \pm m \times x + c \quad (9)$$

$$V_{SD}(t) = -\frac{0,5 \times V_{SD}(t_b)}{T} \times t + V_{SD}(t_b) \quad (10)$$

$$V_{SD,agg}(t) = +\frac{0,5 \times V_{SD,agg}(t_b)}{T} \times t + V_{SD,agg}(t_b) \quad (11)$$

The general exponential function is given by formula 12. The factor c represents the stretching or compression in the x direction. The factor a represents the stretching or compression in the y direction.

$$y(x) = a \times e^{c \times x} \quad (12)$$

" $0 < c < 1$: stretching the graph by a factor " $1/c$ " in x -direction";

" $c > 1$: compression of the graph by the factor " $1/c$ " in x -direction.";

" $0 < a < 1$: compression of the graph by the factor a in " y -direction";

" $a > 1$: stretching of the graph by the factor a in y -direction"

The Nuclis Research Institute finds that aggregated tactical data reaches a certain function value approximately 2.5 times later than the basic function

y(x) (see formula 12). The basic function appears to be stretched by a factor of 2.5 in the direction of the positive x-axis. This results in the value 0.4 for c (see formula 13). In addition, the function starts at time t_b , so the factor a corresponds to the function value at time t_b ($V_{TD_agg}(t_b)$).

Compared to the aggregated tactical data, aggregated operational data reaches a certain function value twice as late. Accordingly, the function is stretched by a factor of 5 in the direction of the positive x-axis. This results in the value 0.2 for c (see formula 14).

$$V_{TD_agg}(t) = V_{TD_agg}(t_b) [10^3 \text{€}] \times e^{\left(\frac{0,4}{1[10^3 \text{Days}]} \times t [10^3 \text{Days}]\right)} \quad (13)$$

$$V_{OD_agg}(t) = V_{OD_agg}(t_b) [10^3 \text{€}] \times e^{\left(\frac{0,2}{1[10^3 \text{Days}]} \times t [10^3 \text{Days}]\right)} \quad (14)$$

For individual tactical and operational data, the progression can be shown on the y-axis, so that the exponent of the exponential function is given a negative sign. Due to the mirrored function progressions, the stretching factor c of the aggregated tactical data corresponds to that of the individual operative data (c for V_{OD} is 0.4). The stretching factor c of the aggregated operational data corresponds to that of the individual tactical data (c for V_{TD} is 0.2). The functions for individual tactical and operational data are shown by the formulas (15) and (16).

$$V_{TD}(t) = V_{TD}(t_b) [10^3 \text{€}] \times e^{-\left(\frac{0,2}{1[10^3 \text{Days}]} \times t [10^3 \text{Days}]\right)} \quad (15)$$

$$V_{OD}(t) = V_{OD}(t_b) [10^3 \text{€}] \times e^{-\left(\frac{0,4}{1[10^3 \text{Days}]} \times t [10^3 \text{Days}]\right)} \quad (16)$$

With the functions shown, it is finally possible to calculate the financial data value of a use case from the end of the observation period t_b to the end of the data lifetime T for different data types.

5. Application in Practice

The previously introduced methodology is applied to three practical use cases in the automotive industry where real-world data is analyzed for its value. The first use case aims at time reduction of failure analysis in vehicle assembly by using production and quality data. The second use case explores the increase of post-purchase sales by using existing vehicle configuration data of existing customers. The third use case addresses the sale of vehicle-generated environmental data to third parties.

For use case 2 some calculations of the data value approaches introduced in Section 4 are presented to illustrate the concept.

The post-purchase upselling use case increases revenues through the sale of additional vehicle equipment based on configuration data of vehicles sold of one vehicle type. Each new vehicle configuration extends the data basis. The data in this use case therefore belongs to the class of aggregated strategic data. With a new series of a vehicle, the equipment packages and configuration options usually change, so that the lifetime of the data corresponds to the lifetime of the series of the vehicle (on average 6.5 years). As the data enables additional sales, it is assigned to the financial benefit category "increase in contribution margin (BCM)". Figure 7 and 8 show some calculations of this use case and summarized results of all use cases in a nutshell.

Data Maturity Level R_{M_total}
$t_2: g_{M1} = \sqrt[3]{g_{M1,C} \times g_{M1,F} \times g_{M1,AC}} = \sqrt[3]{0,25 \times 0,50 \times 0,50} = \mathbf{0,33}$
$t_2: g_{M2} = \sqrt[3]{g_{M2,S} \times g_{M2,5}} = \sqrt[3]{1,00 \times 0,75} = \mathbf{0,86}$
$t_2: g_{M3} = \sqrt[3]{g_{M3,CT} \times g_{M3,A} \times g_{M3,T} \times g_{M3,C}} = \sqrt[3]{0,50 \times 1,00 \times 0,75 \times 0,75} = \mathbf{0,73}$
$t_2: g_{M_total} = \sqrt[3]{g_{M1} \times g_{M2} \times g_{M3}} = \sqrt[3]{0,33 \times 0,86 \times 0,73} = \mathbf{0,48}$
Methodology of cost determination C_{IT}
$FP_{total} = \sum_{n=1}^n FP_n = 2 + 1 + 2 + 2 + 1 + 1 + 1,5 + 1 = \mathbf{11,5 \text{ Function Points}}$
$C_{IT_ext}(FP_{total}) = 0,5 FP_{total}^2 - 6,7 FP_{total} + 35 = 0,5 \times (11,5)^2 - 6,7 \times (11,5) + 35 = \mathbf{24,075 \text{€}}$
Storage Value Potential Methodology SVP
$SVP_D = \left[1 + \left(\sum_{i=1}^n g_{PI_i} \times PI_i \right) + \left(\sum_{e=1}^m g_{PI_e} \times PI_e \right) \right] \times \left(\frac{C_{IT}}{f_q \times g_{M1}} \right) = [1 + (2 + 0,5)] \times \left[\frac{24,075}{0,44 \times \frac{1}{5}} \right] = \mathbf{319,176 \text{€}}$
$SVP_D = \left[1 + \left(\sum_{i=1}^n g_{PI_i} \times PI_i \right) + \left(\sum_{e=1}^m g_{PI_e} \times PI_e \right) \right] \times \left(\frac{C_{IT_total}}{\beta_8 \times g_{M1}} \right) = [1 + (2 + 0,5)] \times \left[\frac{1.000,000}{0,48 \times \frac{1}{5}} \right] = \mathbf{12,152,777 \text{€}}$
Time-dependent economic data value $V_x(t)$
$B_D = (B_{CM,J} - B_{CM}) + (B_{WC,J} - B_{WC}) + (B_{OTEC,J} - B_{OTEC}) + (B_{MC,J} - B_{MC})$
$B_D = (B_{CM,J} - B_{CM}) = (5,600,000 \text{€} - 0 \text{€}) = \mathbf{5,6 \text{ Mio. €}}$
$B_D = 5,600,000 - C_{IT_total} = 5,600,000 \text{€} - 1,000,000 \text{€} = 4,600,000 \text{€}$ where $B_D = V_{SD_agg}(t_b)$
$V_{SD_agg}(t) = \frac{0,5 \times V_{SD_agg}(t_b)}{T} \times t + V_{SD_agg}(t_b)$
$V_{SD_agg}(T) = \frac{0,5 \times 4,6 \text{ Mio. €}}{2000 \text{ days}} \times 2,000 \text{ days} + 4,6 \text{ Mio. €} = 2,3 \text{ Mio. €} + 4,6 \text{ Mio. €} = \mathbf{6,9 \text{ Mio. €}}$

Figure 7: Use Case "Post-Purchase-Upselling" Calculations.

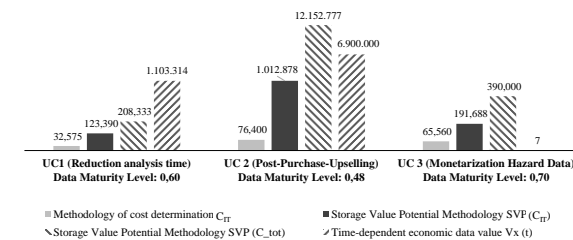


Figure 8. Summary of calculation results for use cases 1 to 3.

6. Conclusions

This paper starts by introducing the theoretical principles and the main factors that influence the value of data. The focus is particularly on the cost types over the data lifecycle, from data collection to data usage.

Subsequently, existing methods for determining data value are investigated by a literature review. Cost, benefit, market price, and risk-oriented approaches are four basic concepts with 22 proposed methods for determining data value. As some of them are not appropriate for the context of automotive manufacturing, a total of twelve methods is selected and investigated in more detail.

In this analysis seven new functional and two new non-functional requirements are derived from the literature review and broken down into three requirement groups. For a cross-application value determination, such methods must include both financial and non-financial evaluation parameters and integrate data quality and cost factors. Furthermore, the four-part data value chain and the time dependency of data value are reflected in the calculation. Moreover, it must be possible to obtain the necessary parameter values at an early stage of the life cycle. A subsequent analysis of the previously described twelve methods coming from the literature shows that none of them comprehensively fulfills all defined requirements. Even if the benefit-oriented methods meet them best, each method of this category fulfills the individual requirements very differently. Consequently, four new methods are developed within this paper, which combine the advantages of existing methods, add new calculation factors, or enhance the given calculation formulas and thereby eliminate the identified inefficiencies of the existing methods.

The newly developed methodology of data maturity assessment as well as data cost assessment allow the prioritization of data business cases in an early phase of the assessment. The data maturity assessment evaluates the quality of the data by a stepwise procedure. The data cost assessment method enables a cost prediction by establishing a mathematical relationship between complexity and cost of data following the function point methodology.

Depending on whether the data is used for operational, strategic, or tactical purposes, different value scenarios are considered. The methodology also looks at the similarity of the use case to be evaluated to other potentially relevant use cases based on the same data. The potential total value of the considered data becomes calculable. The developed methodology of future-oriented data value determination makes it possible to estimate the monetary data value over the whole life cycle.

Three use cases coming from an automobile manufacturing company help to practically validate the applicability of the developed methods with real-world data.

Following the findings of this work, further research is suggested. On the one hand, the quantitative

evaluation of data quality is of crucial importance at all of the developed methods. Since a non-automated quality assessment for large and continuously generated data sets is very time consuming, the methods and algorithms should be implemented as easy-to-use software tools to allow an automated data quality determination based on the quality criteria identified in this paper. On the other hand, there is a need to validate the methods in more practical use cases. The trend curves of strategic, operational, and tactical data shown in Figure 6 must be checked for validity in the context of the automotive industry.

While the methodology developed determines the data value primarily from the internal perspective of an organization, you may also think of approaches that focus on an external perspective. For example, the value of data can be derived by an analysis of the business model of potential buyers of data [23]. Both approaches, the internally and externally oriented view of data value determination, have fundamental synergies, such as the simultaneous consideration of quantitative and qualitative criteria.

7. References

- [1] Fisher, T., *The data asset: How smart companies govern their data for business success*, John Wiley & Sons, 2009.
- [2] Short, J., & Todd, S., *What's Your Data Worth?*, MIT Sloan Management Review, 58(3), 17, 2017.
- [3] Garifova, L. F., *Infonomics and the Value of Information in the Digital Economy*. *Procedia economics and finance*, 23(1), 2015, 738-743.
- [4] DP, K. R. N., *The balanced scorecard. Translating Strategy into Action*, Harvard Business School Press, Boston, 1996.
- [5] Levitin, A. V., & Redman, T. C., *Data as a resource: Properties, implications, and prescriptions*, MIT Sloan Management Review, 40(1), 89, 1998.
- [6] Otto, B., *Quality and value of the data resource in large enterprises*, *Information Systems Management*, 32(3), 2015, 234-251.
- [7] McCall, T., *Understanding the Chief Data Officer Role*, Gartner, Stanford, 2015.
- [8] Sakalaki, M., & Kazi, S., *Valuing and representing information: the paradox of undervaluing information and overvaluing information producers*, *Journal of information science*, 35(2), 2009, 153-164.
- [9] Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B., & Apon, A., *Automotive big data: Applications, workloads and infrastructures*, In 2015 IEEE International Conference on Big Data (Big Data), IEEE, 2015, 1201-1210.
- [10] McKinsey., *Ready for Inspection-The Automotive Aftermarket 2030*, McKinsey & Company, 2018.
- [11] Bertocello, M., Camplone, G., Gao, P., Kaas, H. W., Mohr, D., Möller, T., & Wee, D., *Monetizing car data—new service business opportunities to create new customer benefits*, McKinsey & Company, 2016.

- [12] Laney, D. B., *Infonomics: how to monetize, manage, and measure information as an asset for competitive advantage*, Routledge, 2017.
- [13] Bruendl, S., Matt, C., & Hess, T., Daten als Geschäft—Rollen und Wertschöpfungsstrukturen im deutschen Markt für persönliche Daten, 2016, *Wirtschaftsinformatik & Management*, 8(6), 66-71.
- [14] Spiekermann, M., Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, 54(4), 2019, 208-216.
- [15] Mantena, R., Sankaranarayanan, R., & Viswanathan, S., Platform-based information goods: The economics of exclusivity, *Decision Support Systems*, 50(1), 2010, 79-92.
- [16] Cai, L., & Zhu, Y., The challenges of data quality and data quality assessment in the big data era, *Data science journal*, 14, 2015.
- [17] Gani, A., Siddiqi, A., Shamshirband, S., & Hanum, F., A survey on indexing techniques for big data: taxonomy and performance evaluation, *Knowledge and information systems*, 46(2), 2016, 241-284.
- [18] Redman, T. C., *Data quality: the field guide*, Digital press, 2001.
- [19] Watts, S., Shankaranarayanan, G., & Even, A., Data quality assessment in context: A cognitive perspective, *Decision support systems*, 48(1), 2009, 202-211.
- [20] Janssen, M., van der Voort, H., & Wahyudi, A., Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 2017, 338-345.
- [21] Wang, R. Y., Storey, V. C., & Firth, C. P., A framework for analysis of data quality research, *IEEE transactions on knowledge and data engineering*, 7(4), 1995, 623-640.
- [22] Fisher, C. W., & Kingma, B. R., Criticality of data quality as exemplified in two disasters, *Information & Management*, 39(2), 2001, 109-116.
- [23] Bodendorf, F., Franke, J., A Business Model Analysis for Vehicle Generated Data as a Marketable Product or Service in the Automotive Industry, *Journal On Advances in Systems and Measurements*, 13(1), 2020, pp. 46-55.
- [24] Pipino, L. L., Lee, Y. W., & Wang, R. Y., Data quality assessment. *Communications of the ACM*, 45(4), 2002, 211-218.
- [25] Schmaus, P., Wiesing, B., Baghi, E., & Schmidt, H., Monetäre Bewertung von Stammdaten im Unternehmen, *Controller Magazin*, 41(6), 2016, 10-18.
- [26] Tallon, P. P., Corporate governance of big data: Perspectives on value, risk, and cost, *Computer*, 46(6), 2013, 32-38.
- [27] Sajko, M., Rabuzin, K., & Bača, M., How to calculate information value for effective security risk assessment, *Journal of Information and Organizational Sciences*, 30(2), 2006, 263-278.
- [28] Borek, A., Parlikad, A. K., Woodall, P., & Tomasella, M., A risk based model for quantifying the impact of information quality, *Computers in Industry*, 65(2), 2014, 354-366.
- [29] Glue-Reply: The valuation of data as an asset: a consumption based approach. London, Retrieved from: https://www.reply.com/Documents/13903_img_The-valuation-of-data-as-an-asset.pdf
- [30] Bundesverband Digitale Wirtschaft (BVDW) e.V.: Data Economy: Datenwert-schöpfung und Qualität von Daten. Retrieved from: https://www.bvdw.org/fileadmin/bvdw/upload/publikationen/data_economy/BVDW_Dat enwertschoepfung_2018.pdf, 2018.
- [31] Eppler, M., & Helfert, M., A classification and analysis of data quality costs, In *International Conference on Information Quality*, 2004, 311-325.
- [32] King, D. W., & Griffiths, J. M., Indicators of the Use, Usefulness and Value of Scientific and Technical Information, *Journal of Information Management*, 25(3), 1994, 1-27.
- [33] Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W., A survey on big data market: Pricing, trading and protection. *IEEE Access*, 6, 2018, 15132-15154.
- [34] Leatherberry, T. & Mears, R., Data as an Asset: Balancing the Data Ecosystem. In: *Proceedings of the Fourth MIT Information Quality Industry Symposium*, Massachusetts Institute of Technology. 2010, 348-364.
- [35] BITKOM e.V.: Management von Big-Data-Projekten: Leitfaden, Retrieved from: <https://www.bitkom.org/sites/default/files/file/import/130618-Management-von-Big-Data-Projekten.pdf>, 2013.
- [36] Heinrich, B., & Klier, M., Die Messung der Datenqualität im Controlling, *Controlling & Management*, 53(1), 2009, 34-42.
- [37] Watts, S., Shankaranarayanan, G., & Even, A., Data quality assessment in context: A cognitive perspective, *Decision support systems*, 48(1), 2009, 202-211.
- [38] Trumpetter, J., & Meinken, N., Monetäres Schadenspotenzial von Datenqualitätsfehlern, *Controlling*, 28(10), 2016, 567-576.
- [39] Schmarzo, B., *Determining the economic value of data*. DELL., 2016.
- [40] Bourque, P., & Fairley, R. E., *Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0*. IEEE Computer Society Pres, 2014.
- [41] Sommerville, I., *Software Engineering. Always learning*, 2015.
- [42] Ge, M., Helfert, M., & Jannach, D., Information Quality Assessment: Validating Measurement Dimensions and Process, In: *Proceedings of 19th European Conference on Information Systems (ECIS)*. Helsinki, Finland, 2011.
- [43] Nelson, R. R., IT project management: Infamous failures, classic mistakes, and best practices, *MIS Quarterly executive*, 6(2), 2007.
- [44] Wigand, R. T., Mertens, P., Bodendorf, F., König, W., Picot, A., & Schumann, M., *Introduction to business information systems*, Springer Science & Business Media, 2003.
- [45] Walkerden, F., & Jeffery, R., An empirical study of analogy-based software effort estimation. *Empirical software engineering*, 4(2), 1999, 135-158.
- [46] Nuclis Research Inc., *Guidebook: Measuring the half life of data*. Boston, Retrieved From: <https://nucleusresearch.com/wp-content/uploads/2018/05/m36-Guidebook-Measuring-the-half-life-of-data.pdf>, 2012.