Quantitative Modelling of the Value of Data for Manufacturing SMEs in Smart Service Provision

Jürg Meierhofer¹,* Rodolfo Benedech¹, Lukas Schweiger¹, Cosimo Barbieri² and Mario Rapaccini²

¹School of Engineering, ZHAW Zurich University of Applied Sciences, Switzerland
²School of Engineering, University of Florence

Abstract. The provision of advanced services becomes a relevant differentiation for manufacturing companies, in particular for SMEs (small and medium-sized enterprises). These services, also referred to as smart services, require the collection and processing of data from equipment, customers, and processes, as well as the development of analytics models and the interpretation of their results for improved service value propositions. These steps require significant engagement of the firms in terms of technical and human resources, skills, and new types of value creation processes, which is a major hurdle especially for SMEs. As the value that can be achieved when leveraging the information inherent in the data is not known a priori, the enterprises are not sufficiently informed for taking the decision to engage. Consequently, they are missing out on relevant business opportunities due to a lack of quantitative models for assessing the value of data. In this paper, we discuss the existing literature on data valuation models and explore the state of practice through an interview-based field study. We develop a model for the utility-based valuation of data that helps companies expand their fund of knowledge and skills about the value of their data and thus make better-informed investment decisions. A simulation-based model is developed to support companies in this assessment by providing quantitative insights in the value potential of the data in various use cases. This model opens a series of new research questions for the further elaboration of the data valuation models.

1 Introduction: The Development towards a Data-driven Economy

The shift to services is driven by saturated markets and high competitive intensity [1], as well as by the customer demand for the values and benefits provided by services [2]. In particular, there is an evolution of the customers to demand and pay for some agreed performance output instead of the provider’s resource inputs. Therefore, the transition from goods to services and the addition of services to products is considered essential for manufacturing firms [3]. For the development of the service economy, the omnipresence of

* Corresponding author: juerg.meierhofer@zhaw.ch

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
information and communications technology is a major driving force (Chen et al. 2010). According to [4], the combination of digitalization and servitization lead to a substantial expansion of the service business.

<table>
<thead>
<tr>
<th>Promise to perform outputs</th>
<th>Promise to perform inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase productivity of good</td>
<td>Assist customer in improving his own business processes</td>
</tr>
<tr>
<td>Perform a process on behalf a customer</td>
<td>Facilitate access to and function of the good, before or after sales</td>
</tr>
</tbody>
</table>

**Fig. 1. Input- vs. output-based services (adapted from [2]).**

Based on the value provided to the customer, who is guaranteed either an input or output performance, the literature provides a classification of industrial services [2], [5] (Figure 1). In the columns of the figure, services that are oriented towards the supplier’s goods (left-hand side) or towards the customer’s processes are differentiated (right-hand side). The PLS (product lifecycle services) quadrant contains traditional service models, such as, e.g., the installation of new equipment, maintenance, repair or spare parts delivery. PLS services are complemented or replaced by output-oriented asset efficiency services (AES) when the provider moves to new service models around its products. Examples for this are customization, condition monitoring, predictive maintenance, performance optimization, or consulting for the customer. Such new service models that focus on output performance are also referred to as “advanced services” [3]. For these, the provider guarantees the customer an agreed performance at a given pricing scheme. Depending on the contract, the provider needs to take corrective actions or to pay a penalty, or may have, for instance, a smaller share of revenue, if the agreed performance is not achieved. Along with this, the offering shifts from cost-based to value-based pricing, which gives the provider the chance to generate higher margins if it can lower production costs while maintaining the promised quality of the output. On the other hand, in the case of problems with achieving the agreed performance the provider may encounter financial risks. Assessing these chances and risks quantitatively based on data and analytics is therefore essential for the provider. However, building up the competences and resources to do so requires major investments [6]. Providers are reluctant to take the decision for such investments without knowing the expected benefit quantitatively.

In this context, it is essential for both the provider and the customer to have an estimate of the value created by the services. In the case of advanced services, this value created for the customer directly impacts the willingness to pay and, in return, the value that can be captured by the provider, in other words, the mutual value creation. Therefore, with the increasing degree of servitization of manufacturing and in order to move to advanced services, understanding the value of data for the development and the provision of services...
becomes a key prerequisite, and at the same time it is a key challenge. [7] state that in the data-driven smart manufacturing context data provides the benefits of customization, self-organization, self-execution, or self-learning. This enables data-driven smart services like maintenance, quality control, process monitoring, material logistics, planning, or smart design. Smart service design translates customer voices into product features and quality requirements and thus it accelerates innovation and reduces costs.

According to [6], [8], hurdles for manufacturing companies and, in particular, for SMEs are identified in their lack of hard and soft resources that are required for the provision of data-driven services. One aspect is given by the missing insight in the value that data-driven services can create. In typical cases, there is the chicken-and-egg situation in which the SMEs do not want to invest in leveraging data-driven services until their benefit can be estimated, which is usually not quantifiable before implementing them [9].

Valuation of data for industrial services can also be considered through the lens of Industry 4.0 technology. In general, Industry 4.0 technologies enable companies to offer new services, add functions to their business and turn digitalization into business benefits [10], [11]. Assessing the return on investment for Industry 4.0 enabling technologies is a relevant hurdle for businesses that sell (e.g., the technology provider) or are willing to invest in them (e.g., the customer) [12], particularly in the context of SMEs [13], which often lack analytical data and benchmarks on which to base the estimates of the benefits achievable with the adoption of such technologies and to evaluate their return on investment. Defining which quantitative (e.g., production volumes, costs, quality, and time) and qualitative (e.g., new managing capabilities, business models) value drivers should be considered for the elaboration of their value contribution is problematic since any technology can impact business differently. Once defined, data (e.g., operational, financial) can be used to estimate the return on investment and enable the customer to decide to proceed or not with investment.

Hence, utilizing data in manufacturing and Industry 4.0 contexts is a topic of high importance that requires further steps in research and practical application. In particular, schemes are required that help firms estimate the return value they can expect when investing in data-driven services. This is especially important if firms need to invest in data and service infrastructure before they can leverage the value of data. In that case, having an estimate of an expected return can substantially facilitate the management decision to go further this way.

The research question of this paper is therefore:

**How can firms model the benefit of utilizing data for their service processes and how can they determine this value in a qualitative and quantitative way?**

### 2 Research Methodology

The data valuation framework presented in this paper is elaborated based on a literature review on existing models for the qualitative and quantitative valuation of data with a focus on service value creation (section 3). The service-oriented analysis revealed, among other findings, different perspectives of the data valuation, one of which is the utility-oriented one.

Additionally, an interview-based field study with 8 firms was conducted (section 4). The firms selected for the interviews were either manufacturers being providers or customers of smart products or technology and service providers for such firms. The interviews were conducted with managers responsible for marketing or development of smart products and services. Among other topics covered in the interviews, a strong focus was put on the utility-oriented valuation of data available in the ecosystem.
For this utility-oriented perspective, a quantitative modelling approach was developed taking into account the benefit for different ecosystem actors that can be created by data-driven services (section 5). These quantitative models build on and extend the studies reported in [6], [14]. A demonstrator for the conceptual model was implemented in a simulation model in the simulation tool AnyLogic, which is commercial simulation software suite documented to be well suited for the purpose of hybrid simulation in operational research [15], [16]. This simulation model allowed to run different scenarios of data utilization and to assess their utility.

3 Existing Data Valuation Models in the Literature

From the broad scope of research literature and from practical experiences and case studies documented in the web, it becomes obvious that the positive impact of data on economic value is relevant and undisputed [17], [18]. Data and analytics create relevant new resources that are valuable, rare, costly to imitate, and are supported by and embedded in the organization [19]. When assessing the value of data, a typical scheme consists in considering the difference between the situation with utilizing the data and the one without [20].

Considering data and analytics as drivers for service value creation is framed by the specific literature from the field of Service-Dominant Logic (S-D L) [21]. Both providers and beneficiaries transform data into resources that become cooperative assets and are integrated to new resources [22]. Data can stem from various sources, e.g., equipment sensors, transactions, participation, platforms, interaction with other ecosystems etc. Thus, against the background of S-D L, data as a technological actor can represent an operant resource [23]. Data-driven services provide data in raw or aggregated form to beneficiaries [24]. Given this, the joint sphere in the interaction between providers and beneficiaries [25] is extended and includes more activity and value [24].

The value of data is not determined by itself, but by the value perceived by the customer or beneficiary, as is generally the case with service value [26]. Customer value is a trade-off between benefit and cost, is created in the customer interaction and is contextual and personal. The quantification of the value of data from a service ecosystem perspective and the impact of interaction with humans versus technology (i.e., also data-driven services) represents an open research question [27]. Customer value is multidimensional and consists of multiple types. Based on [28], there are four different and interrelated value dimensions: emotional value, social value (social self-concept), economic value (output-input ratio), and functional value (the utility).

There are different perspectives and models for data valuation in the literature. According to [29], the internet of things and thus data add value by reducing uncertainty and risks. Most sources (e.g., [30]–[32]) differentiate among data valuation models from these three perspectives

a) market value of the data (given by the willingness to pay for a data set)
b) cost-based value (given by the cost to make the data available)
c) utility or functional value (the present value of future utilization of data, e.g., in business processes), which is in-line with the concept of functional value discussed in [28].

According to [32], the valuation based on functional or utility value is conceptually the best approximation, but is difficult to determine and suffers from subjective estimates of experts involved in the estimation. [33] propose a way to make this valuation more transparent and traceable.

Different levels or intensities of value creation with data are reported in the literature. [34] introduce a hierarchy of value creation with smart services that helps to discuss value
creation. Increasing value with smart services and products is created by following these steps: 1) monitoring 2) control 3) optimization, and 4) autonomy. An example for 1) is condition monitoring of machines. The service provider can remotely observe the health condition of the machine running on the customer’s premises. On level 2) control, a feedback loop is established to control the machine based on the outcomes of the monitoring. This may, e.g., result in adapting operational parameters to improve the health condition of the machine. The optimization applied on level 3) pursues a target like, e.g., minimizing energy consumption or maximizing the number of units produced per time. Autonomous systems on level 4) can, e.g., be fully self-organized shop floors. In industrial environments, data is typically used for decision making or decision support [35]. From this perspective, the three steps data – insight – action are described in the literature [24]. The data – information – knowledge – wisdom (DIKW) hierarchy introduced by [36] provides a scheme that helps understanding and communicating the chain of data-driven value creation.

The specific topic of data-based value creation for business process optimization is discussed in many literature sources, e.g., in [11], [17], [37], [38]. A model for the optimization of service value with smart, connected products is elaborated in [39]. According to these, value is created in the domains of, e.g., (health) efficiency and effectiveness, monitoring, quality control, maintenance and support, scheduling, decision making, customer experience and satisfaction, or personalization.

Against the background of the research question of this paper, we focus on the functional or utility valuation of data in the sequel, specifically, the supposedly positive impact of utilizing data for smart services to optimize a customer’s business processes. It is clear that this value creation for the customer in return also results in value capture for the provider, i.e., in mutual value creation in the ecosystem.

4 Field Study

A series of in-depth interviews was conducted with marketing or product innovation representatives of 8 firms. Half of these firms are manufacturers, one of them a customer of smart products and services, the others providers. The other half are technology or service providers for manufacturing firms. The semi-structured interviews were conducted in online video calls and lasted one hour, whereby roughly a third of the time was spent on the companies’ best practices for data valuation. The interviews yielded the results shown in Table 1.

Table 1. Data valuation methods found in the field study.

<table>
<thead>
<tr>
<th>Firm</th>
<th>type of firm</th>
<th>qualitative assumptions of the value of data</th>
<th>quantitative approach to assess value of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>manufacturer (provider)</td>
<td>reduced process costs by less onsite visits, avoiding damages, utilization of data for further purposes, e.g., utilization by other stakeholders</td>
<td>calculation based on process costs, value of the future potential of data today not quantitatively known</td>
</tr>
<tr>
<td>B</td>
<td>manufacturer (customer)</td>
<td>measuring and controlling quality, e.g., by statistical process control and reacting to out-of-band events</td>
<td>calculation of benefit of reduced personnel costs for supervision due to more autonomous machines</td>
</tr>
<tr>
<td>C</td>
<td>manufacturer (provider)</td>
<td>help the customer optimize its fleet, prevent customers from churning, assess customer</td>
<td>calculation of increased customer lifetime value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>manufacturer (provider)</td>
<td>conduct workshops with customers and develop prototype models and pilot trials to quantify the value of the data.</td>
<td>value of the data is determined by the customer’s willingness to pay. The customer itself needs to quantify its willingness to pay by the utility benefit of the data. In order to do so, they explore the data and build first models</td>
</tr>
<tr>
<td>E</td>
<td>technology &amp; service provider</td>
<td>in complex cases, there is a high uncertainty before implementing the data models</td>
<td>estimation based on reference cases which have been solved in comparable industries before</td>
</tr>
<tr>
<td>F</td>
<td>technology &amp; service provider</td>
<td>focus on predictive maintenance: assess the impact on downtime, doing a detail assessment results in too long and costly sales cycles</td>
<td>rough quantification of the economic impact based on a set of questions, simulations if details are required</td>
</tr>
<tr>
<td>G</td>
<td>technology &amp; service provider</td>
<td>chicken and egg problem for utility value of data: value known only after having implemented the service</td>
<td>market value of the data (willingness to pay)</td>
</tr>
<tr>
<td>H</td>
<td>technology &amp; service provider</td>
<td>value of the data given by their utility in processes</td>
<td>calculation based on algorithmic accuracy measures in use cases</td>
</tr>
</tbody>
</table>

We conclude from these interviews that quantitative models for the valuation of data in the manufacturing context are primarily based on its utility value. Only two cases mentioned the market value. Additionally, the analysis shows that firms have a good assumption of the qualitative value of data, but largely lack instruments for its quantification that are practically applicable with reasonably low effort. If quantitative approaches are applied, they require highly specific competences, such as, e.g., for process simulation, and high effort, which cannot be invested before the return on the investment is known (“chicken-and-egg problem”).

5 Findings: Modelling Approach and Demonstrator

5.1 Conceptual Model

The literature review and the field study made apparent that the utility-oriented perspective is the most commonly used by firms and is best suited for the context of value creation by smart services. Utility-oriented or functional service value, as described in [28], [32], assesses the impact of the service in the business processes, e.g., for improving the availability of a resource, its efficiency, or its quality. Given this functional benefit, quantifying the financial value is then straightforward.

Therefore, we developed a model that focuses on the dyadic aspect of a provider-customer relationship of a service ecosystem.
Customers, which in this case are business customers, create data from their operations of products and processes and from individuals. In the manufacturing context, products are typically machines from the provider that are used by the customer for its operations. If these machines are connected in the sense of smart, connected products [34], their sensor generate data which can be shared with the provider, typically over a communications infrastructure such as the internet-of-things (IoT). The same applies to customer processes which are controlled by process management software or workflow tools. Additionally, customer data may originate from individual employees working in the customers’ operations, like data about their actions on the processes and machines. This data can be shared with the provider together with the product and process data. Without going to details, it needs to be mentioned here that sharing data about human individuals requires much more care with respect to data protection and privacy.

If the provider receives this data from the customer, it can create digital models representing the customer’s products, processes, and operations. These models are often referred to as digital twins [40]. The models can be used to improve the value provided to the customer by applying it for services such as, e.g., condition monitoring, machine health prediction, performance optimization, or remote maintenance, which are typically asset efficiency services according to Figure 1.

Given these smart services, the provider creates additional value for the customer by utilizing the data, i.e., by utility or functional value according to section 3. I.e., the provider creates value for the customer in return for the data shared by the customer. As elaborated in sections 2 and 3, the provider can capture part of this value created for the customer, e.g., by an additional willingness to pay for the additional value created by utilizing the data. Additionally, the provider can create value for its own processes based on the data models. E.g., it can reduce logistics costs or assess how its installed base of machines is used by the entirety of customers, thus creating insights for its own marketing or new product development processes.

Overall, the model of Figure 2 shows how sharing data among the provider and the customer results in additional mutual value creation for both of them. This dyadic lens on value creation can be extended to mutual value creation in the ecosystem with a multitude of actors mutually creating value across a multitude of relationships.

5.2 Conceptualization of Demonstrator Model

As mentioned in [6], the hurdles for SMEs to overcome when implementing smart services based on data and analytics are the lack of specific resources and skills. For the management decision whether to invest in such new resources, the SMEs want to estimate...
their future value created by these services, i.e., the value stemming from the data and its processing. The model conceptualized in this paper and its implementation in the demonstrator simulation model are a means to lower this hurdle. Based on the conclusions of [9] and in combination with a field service model of the AnyLogic Library [41], a general demonstrator is created to estimate the advantages of data for an SME with a simulation-model that shows the difference between enabling and disabling data-transfer from provider to beneficiaries. With basic inputs of the SMEs knowledge about their product, services and customers like recommended maintenance period, costs of service or customers’ locations, it is possible to simulate a desired time period and comparing the cases with or without data-transfer. Enabling a data-transfer between provider and beneficiary allows to simulate services like, e.g., remote or condition-based maintenance. Furthermore, advanced knowledge about tools or spare parts needed is possible, so that first level support and initial information collection at the beneficiary’s site is minimized. It is therefore possible to demonstrate the advantages of implementing data-driven services for an SME and estimate a value that it might generate, thus lowering the decision hurdle for the SME to go that direction.

5.3 Implementation of Demonstrator Model

As explained in the methodology section, we conducted interviews with 8 companies. In these interviews, a series of cases for applying data driven services were discussed and identified. Based on these cases, an archetype of a hypothetical SME case was constructed to show the application of the theoretical concept and its implementation in a simulation model.

The hypothetical SME produces machines and is planning to extend its service portfolio with digitalizing its products to offer new, smart services. The provider operates an installed base of machines, which are operated at the premises of the customer (see the example in Figure 3). We assume the provider to have twelve machines in the installed base, one service team and that it expects to assess the return on invest in the smart services infrastructure over a period of 3 years. If the provider has data regarding the machines and service behaviours, as for example the mean time between failure or the location of their machines, the demonstrator can be fed with such data and start calculating.

Fig. 3. GIS-Map with provider (red factory) and customers (cogwheels) in which the service team (lorry) is on the way to a malfunctioning machine (red cogwheel) (hypothetical example).

We model the machine health as a finite state machine [42] by assigning every machine multiple health states such as: working, irregularity, fatal irregularity, maintenance and failed (see the example in Figure 4). These states are all individually modelled by a random process whose parameters are based on the given input by the SME example. If an irregularity leads to maintenance, repair, or replacement, a service team will be sent to the machine to solve the issue.
Fig 4. Downscaled example of a state chart of the machine model.

It is possible to have multiple service teams with different shift models and multiple machines of the installed base. The simulation model is used to run Monte Carlo simulations by executing numerous machine cycles over the period of 3 years, which we consider a time span in the interest of the management. At the end of the simulation runs, resulting values such as the mean performance of a machine, the production quantity, the production loss due to performance loss, costs of service and mean time to repair are provided.

To assess the value of data for smart services, the simulation model allows to enable and disable the utilization of data. If data utilization is enabled, the model incorporates that the provider synchronously gets the information on the current health status of the machine. This is where the benefit of the simulation model comes into play. While implementing this in practice would mean installing sensors in the machine, connecting these to a cloud and giving the provider access to this data, the simulation can simply reflect this by making the variable available to the provider block in the model.

If data utilization is enabled in the model, data-driven services can be applied with the following potential benefits:

- Prevent a failure of the machine by observing a degradation of its health status with some lead time, thus allowing to take measures before a critical state is reached. This condition-based maintenance could of course be further refined into real predictive or prescriptive maintenance. However, for the sake of simplicity of the demonstrator, this is not further elaborated here.
- Enable specific remote maintenance measures such as reboots, software updates, or parameter changes. This has the benefit of shortening the resolution time and saving the costs and energy of the technicians travelling onsite.

For the example case implemented in the demonstrator, Figure 5 shows the comparison of the simulation runs over 3 years for the case with utilizing the data (left hand side) and the one without (right hand side). The case with the data utilization enabled shows that the average machine performance turns out to be higher – 93.19% vs. 85.08% - and the average time to repair shorter – 9.08 days vs. 16.29 days. Utilizing the data for the services has thus the economic benefits of lower production loss due to avoided or shorter break down times and lower service costs due to more efficient remote services, besides the very important non-financial benefit of improved customer and employee experience as well as positive impacts on the reputation of the provider and the customer (with these non-financial benefits not being quantified by this model).
Fig. 5. Comparison of enabled (left) and disabled (right) data-sharing of a single simulation run in the demonstrator. The time stack chart on top shows the monthly averages of states of all machines of the installed base, the line chart in the middle shows the weekly average performance and the line chart at the bottom shows the monthly average time to repair or downtime of machines.

6 Discussion and Outlook

In this paper, we showed by a literature analysis that there are different perspectives on the valuation of data and that the utility or functional value perspective lends itself best for a service-oriented modelling. Through the service lens, utilizing data has the potential to increase the value created in a qualitative and quantitative way. The field study has revealed that most companies still lack theoretical and practical tools for quantifying the value of the data in their ecosystem. This represents a hurdle for investing in the infrastructure required for data-driven (also known as smart) services. For an informed investment decision, companies need to be enabled to quantify the potential return on their investment. To circumvent this hurdle, companies often start with small and relatively low-cost pilot projects revealing parts of this value or by roughly estimating the value of the data by comparable best practice cases.

The concept discussed in this paper models the value of data for service provision in manufacturing ecosystems quantitatively by taking into account the impact on service process if the data available in the ecosystem is shared and used. Using the data has potential positive impact on the output performance of machines by preventing degradations, reducing the time to react, or reducing the effort to react (e.g., by remote
services or customer self service). Additionally, the provider can potentially create value from the data for its marketing or new product development process. Overall, this results in increased mutual value creation by the ecosystem actors integrating the available data and knowledge resources to create new resources.

Referring to the research question formulated in section 1, the conceptual model and the simulation model can now be applied for specific SME cases to assess the value of the data in their ecosystem. To do so, the service processes are modelled a) for the situation without utilizing the data and b) for the one with the data being utilized. Comparing the resulting performance variables provides the expected benefit of the data. In this modelling process, simplifying assumptions need to be made for translating the availability of data into indications for the relevant conditions of the machine. This explicitly neglects that different quality of data may result in different quality of the condition indication, which is subject to future research.

Future work will expand the model with new functions, which will allow to map companies and their services more accurately by enabling and disabling specific services which are appropriate for a company’s business model. The current model focuses on the use phase of the customer lifecycle. By an expansion on the entire customer lifecycle perspective, positive impacts of utilizing data on marketing and sales as well as customer retention can be added [39]. This will take into account new elements such as customer behaviour based on marketing measures and customer loyalty measures. Additionally, future work will apply and parametrize the model for specific company cases, thus enabling to develop industry-specific blueprints of the model.

There are a couple of limitations and open research questions in this study. Further research should, among others, also address the following issues:

- How to operationalize the application of the model in a way that allows to adapt it for a new SME case in a low effort yet systematic process, thus serving as an adequate support tool for the investment decision?
- How to extend the value creation model beyond the dyadic relationship between provider and beneficiary towards a more ecosystem-oriented view?
- How to incorporate a quantification of value for non-functional and non-financial dimensions?
- How to incorporate the quality of the data in way that keeps the model application simple enough for supporting investment decisions in SMEs?
- How to take into account the different accuracies of different data science modelling methods and the cost to achieve these levels of accuracy?

Acknowledgments

The authors would like to thank ZHAW digital for the support of this work. This study was also partly supported by the grant "ABH097" within the framework of the Interreg VI-programme "Alpenrhein-Bodensee-Hochrhein" (DE/AT/CH/LI) whose funds are provided by the European Regional Development Fund (ERDF) and the Swiss Confederation. The funders had no role in study design, data collection and analysis, decisions to publish, or preparation of the manuscript.

References


