



# Generating Customer Insights Using the Digital Shadow of the Customer

Kristof Briele<sup>1</sup>(✉), Marie Lindemann<sup>1,2</sup>, Raphael Kiesel<sup>1,2</sup>, and Robert H. Schmitt<sup>1,2</sup>

<sup>1</sup> Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany  
k.briele@wzl.rwth-aachen.de

<sup>2</sup> Fraunhofer Institute for Production Technology IPT, Steinbachstr. 17, 52074 Aachen, Germany

**Abstract.** Smart products, Social Media and innovative market research lead to an abundance of customer data, yet due to their heterogeneous sources and structures, they are scattered throughout the company. Joining these different types of data can lead to a large gain in customer insights that would not have been possible by analyzing the data individually. It is a necessary step for the transition of the current mostly hypothesis-based product design process towards a data-driven one and enables accelerated product development with truly innovative products tailored to the customer. This paper explains the holistic approach to identifying customer needs and requirements: the digital shadow of the customer. It is a concept transferred from the Internet of Production and its digital shadows of products and processes. The paper first gives an overview of customer data that form the customer data lake and reviews current data analysis methods using an explorative literature review. We then explain the concepts of the digital shadow and data lake, their main principles and benefits of using digital shadows for product development.

**Keywords:** Product development · Customer data · Internet of Production

## 1 Motivation

The widely accepted understanding of Industry 4.0 and the increasing digitalization is that it will lead to better, faster and/or more robust decisions. This development can be traced back to early statistical analysis in quality management but is currently represented in real-time models of highly complex production processes like fine blanking or die-casting [1]. The *Internet of Production* is a framework for Industry 4.0 to aggregate data, derive digital representations such as *Digital Shadows*, and support the decision-making process [2]. It includes product development as one of its three phases with the goal of transforming hypothesis-based product design into a customer-centered and data-driven one. However, customer data are more subjective than production data and the decisions have a higher impact on the overall product quality [3, 4].

This research starts with a comprehensive overview of customer data and the corresponding analytical methods through an explorative literature review and subsequent

derivation of their properties. The analysis of both data and analysis methods leads to a discussion of if and how the concepts of the Internet of Production can be applied to customer data and product development.

## 2 Analysis of Customer Data for Product Development

In product development, companies can use a variety of customer data to derive customer insights and improve the product. However, research focuses on the use of individual data types. The goal of this chapter is to achieve a basic understanding of the relevant data and their properties. It is a necessary step for automating the data analysis in the future.

### 2.1 Data for Product Development

Lindemann et al. proposed a first overview of customer data types relevant for product development including *social media data*, *complaint data*, *user opinion survey*, *lead user workshops*, *expert interviews*, *internal audits*, and *descriptive studies* [5]. The first part of our research aims to create a more holistic overview by both extending and validating the aforementioned data types. This is achieved by conducting an explorative literature review that focuses on deriving a broad overview of the data types and not a structured, quantitative analysis of the data. The literature review is combined with a brainstorming workshop with five experts who are experienced in both research and industry in the field of customer insights and product development of consumer goods; their input extended the list of relevant data types. In the literature review, redundant data types do not affect the importance and benefits of each individual data type and are, therefore, not registered. Keywords of this review are: *customer data types*, *customer data sources*, *product development data*, *product design data*, *customer profile*, *customer data categories*, *customer data properties*, and *voice of the customer*. The publications include journals and conference proceedings published from 2015–2021, but some older publications are also accepted. From a total of 53 papers, depending on the manner of counting, approx. 100 different customer data examples are extracted. To achieve a comprehensive overview, the expert workshop identified eight main customer data types from the full list used in product development (see Table 1). Data types and data sources are strongly connected and the focus of this paper is not to show a definite separation, thus, we use the term data type going forward. Twenty-seven exemplary literature references are listed in the table that represent the spectrum of the customer data examples.

**Table 1.** Relevant customer data types for product development

Data types	Examples	Literature
Social Media Data	Customer reviews, discussion forums	[6–11]
Complaint Data	Formal complaints	[12]
Customer Study Data	Interviews, questionnaires, behaviour analysis, lab experiments, workshops	[12–19]

(continued)

**Table 1.** (continued)

Data types	Examples	Literature
Usage Data	Visits to websites, click stream, customer location, banking logs	[13, 18, 20–22]
User Data	Customer type, demographics, 3D anthropometry	[23–27]
Sales Data	Purchase history, behavioural analysis	[20, 23]
CRM Data	Relationship details, call history, interaction data	[23, 28]
Market Data	Neuromarketing, crowdsourcing, market surveys	[25, 29–32]

Although all examples fit into these main data types, a clear assignment to one data type and a distinction between data type and data source may be difficult. E.g., the difference between social media data and complaint data is blurry on the content level since social media data contain complains; yet these data types differ strongly in their data structure as social media data are more unstructured texts rather than complaints though a standardized form.

Standardized descriptions of data using properties such as “structure” are helpful to make use of this heterogeneous pool of customer data. With these descriptions, companies can more easily decide which data types to use to reach a certain decision. Furthermore, the description is a necessary step for automated data analysis. Lindemann et al. proposed a first set of properties of customer data, which we extended to include *subjectivity*, *degree of structure*, *degree of specificity*, *number of data points*, *update frequency*, and *cost* [5].

In product development, not only customer information is relevant but also information about the product and process capabilities; thus, the customer data are extended by product data. Since the focus of this research is on customer data, the list of product data types are not derived by a literature analysis but in a second workshop with the same five experts to get a first understanding of product data. The experts were given the previously accumulated list of data types and assessed both, product and customer centered data types, using the accumulated knowledge of the literature review and their experience (see Table 2).

It shows that product and customer data differ – as expected – in *subjectivity*. Design parameters, e.g., are precise values while measurements only vary inside the accuracy of the measurement system. Customer data, on the other hand, include the subjective opinion of different customer groups due to factors such as age or cultural background. In general, this high *subjectivity* is accompanied by a lower *specificity*. The main reason is that a high specificity requires a defined research question, which limits the expression of subjective opinions. The structure of the data is subject to similar effects. A high degree of *structure* limits the customer’s freedom of expression but also limits the subjectivity. An equally high degree of structure, specificity and subjectivity – as represented by customer study data – is expensive to buy, because it requires a dedicated team investigating a limited number of research questions.

An interesting development in customer data is the integration of “newer” data such as *social media* and *usage data* from automated sensors in cyber physical systems (e.g. smart fridges). Both offer a vast amount of data that provide insights into unfiltered customer opinions and real-life usage behavior. The necessary amount of data requires automated data crawling and analysis to be financially valuable.

**Table 2.** Structured data types for product design assessed in 6 properties

		Subjectivity	Degree of Structure	Degree of Specificity	Number of Data Points	Update Frequency	Cost
Customer Centered	Social Media Data	●	○	◐	●	●	◐
	Complaint Data	◐	◐	◐	◐	●	◐
	Customer Study Data	●	●	●	◐	◐	●
	Usage Data	◐	●	◐	●	●	◐
	User Data	◐	●	○	◐	◐	◐
	Sales Data	◐	●	◐	◐	◐	◐
	CRM Data	◐	◐	○	◐	◐	◐
	Market Data	◐	◐	○	◐	◐	◐
Product Centered	Quality Control Data	○	●	●	◐	◐	◐
	Measurement Study Data	○	●	●	◐	◐	●
	Design Parameters	○	●	●	◐	◐	◐
	Manufacturing Processes Data	○	◐	◐	●	●	◐
	Laws and Norms	○	◐	●	◐	◐	◐

low ○ ◐ ◑ ◒ ◓ high

### 2.2 Analysis Methods and Goals

Companies use different methods to acquire and gain insights from customer data (*customer insights*). These methods differ in terms of the research question, the scope of the method and the output. Simplified, they are divided into two steps: First, the company addresses the question of *which* features determine the customer experience, and consequently, *how* these features should be specified to enhance the product experience.

The first step toward obtaining an overview of product acceptance on the market is opinion mining on the basis of customer reviews [33]. The output includes terms that describe or evaluate the product and their frequency. This can help determine the customers’ impressions of the product, their overall satisfaction, and establish the connection between the emotions and sentiments about certain product features. An alternative approach is eye tracking. Heat map analyses of the recorded eye movement data when

observing a product can be used to determine focus points and gaze time of the customer when participating in a customer study [34]. These results can be used to draw conclusions about the perceived quality of the product. Other methods for determining what drives the customer experience can be the Kano model [35] or Failure Mode and Effects Analysis (FMEA).

In the second step, individual product features are analyzed. Kansei engineering aims to develop or improve products by translating the customer's psychological feelings and needs into parameters [36]. The output includes quantitative links made between so called Kansei words and product features. As a result, products can be designed to evoke the intended feeling. Another method to learn more about the customer's desired and undesired characteristics of a product is Conjoint Analysis. This method estimates the structure of consumers' preferences by using their overall judgements about a set of alternatives, specified by expressions of different features [37]. Other methods are VR Testing, Total Quality Management or User Centered Design [38].

The aim of this multi-step approach is to combine the results of the individual methods to obtain a larger picture of customer requirements and how these can be fulfilled. The gained insights must then be integrated into the product development process and the company's knowledge management.

### 3 Enabling Customer-Centered Data-Driven Product Development

The Internet of Production as a reference infrastructure for Industry 4.0 offers two concepts of particular interest for the storage, analysis and application of data: the *data lake* and the *digital shadow*. Both concepts are researched in depth, especially machine data and the subsequent production planning. [2].

We discuss both concepts for the application in product development. The addition of customer data is an especially novel but important step for data-driven decisions in every step of production. After a brief general explanation, we lay out potentials and requirements based on the previous chapter about customer data. This is the second contribution of this paper since the discussed concepts do not include customer data so far.

#### 3.1 The Data Lake

The data lake is the foundation of the Internet of Production to realize a fully interconnected production landscape. In contrast to other data warehouses, it allows for storing raw data in an unstructured "storage first, query later" manner. This approach offers real-time control of tightly integrated production processes, storing and processing heterogeneous production data, and secure privacy-aware collaboration [1].

Customer data in product development is an even more heterogeneous pool of data that ranges from unstructured opinions of a single customer up to broad market analyses. This makes data and knowledge management a key challenge in product development [39]. The data lake is not a rigid construct and is explicitly suitable for unstructured data. It can serve as a database for both customer wishes and production capabilities.

Extending the data lake to include customer data may improve the data availability and accessibility across the company. The decisions in product development are typically made in weeks rather than milliseconds as they serve to control production machines. Different departments, however, require access to the data at different phases of product development. Since the data lake stores the entire data history with high accessibility, the requirements of customer data are met. It offers the possibility to access historical customer information and decisions, which do not change as often as, e.g., innovations, and make decisions based on transparent data more robust against budget cuts.

The data lake is designed for large data input streams from sensors of production machinery in the magnitude of 6.2 Gbit/s [1]. For customer data, especially for usage and social media data, high data streams are also important. E.g., a smart fridge with a small uplink of only 1 kB/s accumulates up to 1 GB/s in total for a company that has sold one million fridges. Though the data do not need real-time analysis per se, the constant data stream of highly distributed units must be considered.

### 3.2 The Digital Shadow

The digital shadow is a “*sufficiently accurate mapping of the processes in production, development and adjacent areas with the purpose of creating a real-time capable evaluation of all relevant data*” [40]. Since most digital shadows describe production machines, Gussen et al. created the first definition of the digital shadow of the customer as all “*data that encompass aspects of the direct interaction between the customer and the product*” [41].

Using the overview of data and data analysis methods in Sect. 2, we can further refine the understanding of the digital shadow. In product development, the digital shadow of the customer is used together with information about product and process capabilities to create a holistic base of information. The key challenge is to combine different data types to derive customer insights that are feasible in the production landscape. Meyer et al. propose the idea of redefining the digital shadow during the product development process [42]. The sequential methods regarding the questions: *which* product features have an impact on the customer and *how* these features have to be specified, fit into this understanding of refinement.

Not only data types and data analysis methods are heterogeneous, but the resulting information is also used by various departments with a specific purpose at different phases of the development process. The digital shadow must be task specific. In product development, objectives vary from increasing the Perceived Quality, reducing complaints, generating innovations or tailoring the product experience. The digital shadow must allow for this range of tasks by providing relevant information for specific questions.

In summary, both concepts, data lake and digital shadow, can be used for customer data and product development. However, the decisions in product development are more diverse and are based on multiple data types. An automation of the data processing and analysis – the vision of the Internet of Production – requires a deep understanding of the data’s structure and application possibilities.

## 4 Conclusion

The research shows the heterogeneity of data used for product development and their application in a company. An explorative literature review and two expert workshops identified eight customer data and five product data types relevant for product development, which we assessed based on five main properties. The analysis of data types and analysis methods in the context of the Industry 4.0 framework, Internet of Production, showed that its main principles, the data lake and digital shadow, are applicable to customer data and have the potential to increase automation of analysis and decision quality. Further research is necessary on integrating the data lake into a company's knowledge management and further describing data types as machine-readable input using, e.g., UML.

**Acknowledgement.** Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

## References

1. Pennekamp, J., et al.: Towards an Infrastructure enabling the internet of production. In: 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)
2. Brecher, C., Brockmann, M.V.: Produktion durch Digitale Schatten – Werkzeugmaschine 4.0, S 543–552
3. Stylidis, K., Wickman, C., Söderberg, R.: Perceived quality of products: a framework and attributes ranking method. *J. Eng. Des.* **31**(1), 37–67 (2020)
4. Krishnan, V., Ulrich, K.T.: Product development decisions: a review of the literature. *Manage. Sci.* **47**(1), 1–21 (2001). <https://doi.org/10.1287/mnsc.47.1.1.10668>
5. Lindemann, M., Nuy, L., Briele, K., Schmitt, R.: Methodical data-driven integration of perceived quality into the product development process. *Procedia CIRP* **84**, 406–411 (2019)
6. Mudambi, S., Schuff, D.: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com (2010)
7. Nambisan, S., Baron, R.A.: Organizing virtual customer environments to promote two types of customer contributions. *Organ. Sci.* **21**(2) (2010)
8. Bae, J.K., Kim, J.: Product development with data mining techniques: a case on design of digital camera. *Expert Syst. Appl.* **38**(8), 9274–9280 (2011)
9. Greco, F., Polli, A.: Emotional text mining: customer profiling in brand management. *Int. J. Inf. Manage.* **51**, 101934 (2020)
10. Jin, J., et al.: Review on recent advances in information mining from big consumer opinion data for product design. *J. Comput. Inf. Sci. Eng.* **19** (2019)
11. Sandor, L., et al.: Social media in product development 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, S 88–93 (2018)
12. Birch-Jensen, A., Gremyr, I., Halldórsson, Á.: Digitally connected services: improvements through customer-initiated feedback. *Europ. Manage. J.* **38**(5) (2020)
13. Wedel, M., Kannan, P.K.: Marketing analytics for data-rich environments. *J. Mark.* **80**(6), 97–121 (2016). <https://doi.org/10.1509/jm.15.0413>
14. Teixeira, T.S., Wedel, M., Pieters, R.: Moment-to-Moment optimal branding in TV commercials: preventing avoidance by pulsing. *Mark. Sci.* **29**(5), 783–804 (2010)

15. Schöggel, J.-P., Baumgartner, R.J., Hofer, D.: Improving sustainability performance in early phases of product design: a checklist for sustainable product development tested in the automotive industry. *J. Clean. Prod.* **140**, 1602–1617 (2017)
16. Guo, F., et al.: Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered. *Int. J. Ind. Ergon.* (2016)
17. Lee, J.-H., Chang, M.-L.: Stimulating designers' creativity based on a creative evolutionary system and collective intelligence in product design. *Int. J. Ind. Ergon.* **40**(3), 295–305 (2010). <https://doi.org/10.1016/j.ergon.2009.11.001>
18. Blocker, C.P., Flint, D.J., Myers, M.B., Slater, S.F.: Proactive customer orientation and its role for creating customer value in global markets. *J. Acad. Mark. Sci.* **39**(2) (2011)
19. Wall, M., Ulrich, K., Flowers, W.: Evaluating prototyping technologies for product design. *Res. Eng. Design* **3**, 163–177 (1992)
20. Greenberg, P.: The impact of CRM 2.0 on customer insight. *Jnl. Bus. Indus. Market.* **25**(6), 410–419. <https://doi.org/10.1108/08858621011066008>
21. Hossain, M.A., Akter, S., Yanamandram, V.: Revisiting customer analytics capability for data-driven retailing. *J. Retail. Consum. Serv.* **56**, 102187 (2020)
22. Balomenou, N., Garrod, B.: Using volunteer-employed photography to inform tourism planning decisions: a study of St David's Peninsula, Wales. *Tourism Manage.* **44** (2014)
23. Earley, S.: The role of a customer data platform. *IT Prof.* **20**(1), 69–76 (2018)
24. Verwulgen, S., et al.: A new data structure and workflow for using 3D anthropometry in the design of wearable products. *Int. J. Ind. Ergon.* (2018)
25. Ho-Dac, N.N.: The value of online user generated content. *J. Bus. Res.* **112**, 136–146 (2020)
26. Petukhov, A., Zaikin, O., Bochenina, K.: Analysis of the geospatial activity profiles of bank customers. *Procedia Comput. Sci.* **156**, 245–254 (2019)
27. Berry, M.J.A., Linoff, G.: Data mining techniques. For marketing, sales, and customer relationship management, 2. Aufl. Wiley Pub, Indianapolis Ind. (2004)
28. Tang, A.Y.C., et al.: Application of data mining techniques in customer relationship management for an automobile company ICIMU (2004)
29. Mansor, A.A., Mohd Isa, S.: Fundamentals of neuromarketing: what is it all about? *Neurosci. Res. Notes* **3**(4), 22–28 (2020). <https://doi.org/10.31117/neuroscirn.v3i4.58>
30. Poetz, M.K., Schreier, M.: The value of crowdsourcing: can users really compete with professionals in generating new product ideas? *J. Prod. Innov. Manage.* **29**(2), 245–256 (2012)
31. Albers, A.A., et al.: requirement profiles as a foundation for customer-centric ideation in business-to-business markets. In: *Procedia CIRP* (2019)
32. Hasoneh, A.: Measuring the effectiveness of e-mail direct marketing in building customer relationship. *Int. J. Market. Stud.* (2010)
33. Yang, B., Liu, Y., Liang, Y., Tang, M.: Exploiting user experience from online customer reviews for product design. *Int. J. Inf. Manage.* **46**, 173–186 (2019)
34. Liu, C., et al.: Application of Gaze Heat Maps in Impression Evaluation. *IJAE* (2019)
35. Sharif Ullah, A.M.M., Tamaki, J'i.: Analysis of Kano-model-based customer needs for product development. *Syst. Engin.* **14**(2), 154–172. <https://doi.org/10.1002/sys.20168> (2011)
36. Japan Society of Kansei Engineering *Kansei Engineering International Journal*
37. Getting Started with Conjoint Analysis Strategies for Product Design and Pricing Research
38. Zöllner, S.: Mapping Individual Subjective Values to Product Design
39. Holmlund, M., et al.: Customer experience management in the age of big data analytics: a strategic framework. *J. Bus. Res.* **116**, 356–365 (2020)
40. RWTH Aachen Campus Production Engineering (2017) Digital Connected Production
41. Gussen, L.C., et al.: Your customer's digital shadow – objectification of human perception by a multisensory measurement system. In: *Procedia CIRP* (2020)
42. Meyer, R., et al.: A Tangible XR for Rapid Prototype Development, Design Exploration and Usability Assessment. *AHFE* (2021)