Valid Customer Data: The Foundation for Omni-channel Marketing

Simone Braun and Andreas Heißler

1 Digital Transformation Sets Customer Data into Focus

The digital transformation continues unabated and is bringing about a profound change in the economy, society, and politics. Information and communication technologies are comprehensively penetrating and connecting all areas of the economy and our daily lives (BMWi, 2017). This, together with developments such as artificial intelligence, cloud computing, or the Internet of Things, provides new innovative value creation opportunities for companies. To that end, data forms the foundation. Data has become an essential asset for new and competitive services, customer access, business models, and innovations (BVDW, 2018; Azkan et al., 2019).

A study by the European Commission revealed a market of 400 billion euros for the European data economy in 2019 for the EU27 plus the United Kingdom—a growth of 7.6% compared with the previous year. According to estimates, the European data economy (EU27 excluding the U.K.) is expected to reach a size of between 432 billion euros and 827 billion euros by 2025 (European Union, 2020). However, companies in Germany have only just started to exploit the potential of data. There are major weaknesses in the areas of data management and data governance, including data quality (DEMAND, 2019).

Communicating via social media platforms, managing finances from a smartphone, and shopping online at any time and anywhere have become a matter of course for many people. Users expect customized services in return for their

S. Braun (✉)
Offenburg University of Applied Sciences, Offenburg, Germany
e-mail: simone.braun@hs-offenburg.de

A. Heißler
Uniserv GmbH, Pforzheim, Germany
e-mail: andreas.heissler@uniserv.com

U. Hannig, U. Seebacher (eds.), *Marketing and Sales Automation*, Management for Professionals, https://doi.org/10.1007/978-3-031-20040-3_10

personal data. Kreutzer & Land (2013, p.47) characterize today's expectations on the customer side with the catchwords "I, everything, immediately and everywhere," with the smartphone becoming the personal communication center.

The buying behavior of consumers has changed dramatically. Today, prospects search, evaluate, and decide in a different way. Numerous studies show that recommendations from family and private contacts, e.g., via social media, and increasingly also by influencers, influence consumers' purchasing decisions (Bognar et al., 2019; Gafni & Golan, 2016; SPLENDID RESEARCH, 2018; Greven Medien & GfK, 2019).

More than half of online shoppers use customer reviews as decision support before making a purchase (Bitkom, 2020). This change does not only affect the Business-to-Consumers (B2C) sector but also Business-to-Business (B2B) procurement; the decision-makers transfer their private consumption and communication behavior to their professional contexts (van Bommel et al., 2014). Research, opinion-forming, and supplier selection increasingly take place online. More than half of the B2B purchasing decision process is already completed before contact is made with a supplier under consideration for the first time (Maycock et al., 2012; Google, 2016).

Thanks to modern Big Data methods, it is now easier than ever to collect data. This is especially true for customer data. Today, every contact a person has with a company is meticulously and precisely recorded. Marketers can read from the data, for example, how their customers are responding to the currently running campaign.

Digital marketing professionals have high expectations in artificial intelligence (AI), machine learning, and predictive analytics. According to the Digital Dialog Insights study, nine out of ten professionals surveyed see AI as a promising technology, particularly for personalization (Eichsteller & Seitz, 2019). The first approaches are already in use, e.g., to analyze customer data and recognize patterns from which customer segments are formed, suitable content is generated, and automatically displayed (Gentsch, 2017; Kröger & Marx, 2020). AI-based processes increasingly gain popularity, e.g., in e-commerce for dynamic pricing, customer dialog via chatbots, or personalized product recommendations (Bernhard & Mühling, 2020). Thus, B2C marketing decision-makers see data and analytics as a top priority (Myler, 2017).

But correct predictions of the next best action based on an immeasurable amount of data require clarity about what data is needed and whether it is of sufficient quality. AI-based approaches can only bring the best benefits if the data they use for learning is of high quality (Stern et al., 2021; Stevens, 2019; Weber et al., 2021). Any algorithm that is trained on incorrect data will come to wrong conclusions (Kessler & Gomez, 2020). For example, if customer data was collected with a wrong salutation, an online shop would, in simplified terms, recognize an increased interest in women's handbags for men—with corresponding consequences for the display of automated recommendations. If the date of birth is incorrectly recorded, a 35-year-old customer, for example, would be sorted into the 60+ age group and addressed as a sprightly senior citizen. If the date of birth is missing altogether, the customer might not be addressed at all (Wolters, 2020). Although there are certain data pre-processing techniques for AI-based methods to deal with missing values (i.e., imputation methods), these can only be used if missing values are unbiased, and a sufficiently large amount of data is available (Berthold et al., 2020). However, if there is a systematic error, e.g., the birth dates of a certain customer group are missing, this can lead to distorted results of the learning algorithms.

Besides ensuring a high quality of customer and prospect data, all such available data needs to be consolidated and made available to the AI system. However, the quality and completeness of data is the major challenge in marketing—a fact that marketers are well aware of (Eichsteller & Seitz, 2019; Lünendonk, 2019; Myler, 2017; Data IQ, 2017). This is also confirmed by a study commissioned by Uniserv in spring 2020[1] among decision-makers on the status of customer data management in corporations (Uniserv, 2020). Almost 15% of the participants stated that a lack of data quality leads to difficulties in projects involving artificial intelligence. Accordingly, 82% of respondents indicated that consistently high data quality is important. Seventy percent see a 360° customer view as important. In a study commissioned by Uniserv from 2019,[2] 51% of respondents also stated that there is a close connection between data quality and the success of AI-related projects (Uniserv, 2019).

For this reason, after initial euphoria, disillusionment spreads because desired results are not achieved. To get a 360° customer view in marketing automation and to develop a personalized cross-channel marketing strategy, reliable master, transaction, and behavioral data is of utmost importance. Data quality is not only essential for the relevance of marketing campaigns, but also for customer satisfaction. The following chapter gives an overview and understanding of different types of customer data.

2 Customer Data

Customer data can be divided into fixed data—master data or core data—and variable data—transaction data. The latter being one of the most important data assets, especially for analysis.

Definition of Master Data According to ISO 8000, master data is defined as entities "which are independent and fundamental to an organization; which must be referenced in transactions in order to perform them" (ISO, 2016).

Master data describes a company's core operational objects, also known as business objects, such as customers, business partners, suppliers, products, or

[1] Cf. Customer Data Management 2020 Practice Study, Uniserv GmbH/YouGov, 204 executives from German companies, survey May 22 to June 02, 2020, anonymous online survey.

[2] Cf. Trend Study Customer Data Management 2019, Uniserv GmbH/Grohmann Business Consulting, 154 participants from German companies, survey December 2019 to April 2019, anonymous online survey.

personnel (Piro & Gebauer, 2021). Master data is fundamental for ongoing business operations and is (semi-)static, i.e., it usually changes rarely. Personal master data includes, for example:

- The full name
- Title
- Date and place of birth
- Gender
- Residential, shipping, billing addresses
- Contact information, such as phone numbers, mobile numbers, and e-mail addresses
- Personal identification numbers, such as social security, passport, driver's license, tax
- Login/account data, such as bank details, loyalty card, Twitter handle, LinkedIn address, or company-specific IDs

Master data can be supplemented by other sociodemographic attributes that give a more comprehensive picture of the customers. These attributes vary widely between companies and include, for example, information about:

- Family such as marital status or number of children
- Professional history and related attributes such as income, industry, qualification, or position
- Lifestyle such as living situation, type of vehicle or pets
- Hobbies such as memberships in clubs, gyms, or subscriptions
- Special categories of personal data[3] such as
  - Ethnic origin
  - Political opinions
  - Religious or ideological beliefs
  - Union affiliation
  - Genetic data
  - Biometric data for the unique identification of a natural person
  - Health data
  - Data on sex life or sexual orientation

Definition Transaction data Transaction data is transaction-oriented and provides information about activities and individual transactions of the core business objects of a company (Otto & Österle, 2016).

[3] These data are particularly data worthy of protection. Their processing is prohibited (Article 9 (1) of the General Data Protection Regulation). Exceptions are, for example, the explicit consent of the data subject (Art. 9 (2) (a)) or medical diagnostics (Art. 9 (2) (h)).

Transaction data is created repeatedly, changes frequently in operational processes, and contains references to master data. Examples include invoices, orders, deliveries, returns, etc., with information on the number, type, time, and price of purchased or returned items, order and payment data, and others.

In addition to master data and transaction data, behavioral or user interaction data play an important role for marketing, retail and e-commerce and should be distinguished as a further group at this point.

Definition of behavioral data Behavioral data is any data about the behavior of an individual person that is collected or derived, especially, for marketing purposes.

This can be information on communication and purchasing behavior, brand preferences, or product usage. This category also includes information about:

- Online activities and interactions with online content such as website visits, click streams, retention time, product views or social media engagement; collected, for example, by cookies, session IDs, etc.,
- E-mail communications, such as openings, click-throughs, or replies to messages, or
- Customer service interactions, such as details of inquiries, communication time, details recorded by service personnel, and
- End devices used, IP addresses, and geolocations.

Behavioral data enables companies to understand how individual customers interact with a company, whether through specific actions or reactions. For example, the recency, frequency, monetary value (RFM) analysis (Bult & Wansbeek, 1995) is used to score a customer's value. Applied on transaction data, this can disclose information about how recently and how often purchases are made and how much is spent, and, derived from this, the preferences of an individual.

From a structural point of view, customer data can be divided into three categories: structured, unstructured, or semi-structured. "Structured data is data for which structuring information—metadata—is available" (Piro & Gebauer, 2021, p. 146). These are, for example, the format and permitted values of a date of birth data attribute. On the other hand, data from text, audio, video, and sensors exist in a non-formalized structure and is, therefore, referred to as unstructured. Semi-structured data is a hybrid of structured and unstructured data. That is, certain parts may have structure, but overall, there is no specific and unique structure. For example, an e-mail contains structured information about who sent it and who received it. However, the content of the e-mail, as a written message, is largely unstructured. Other examples include voice and text input when interacting with a chatbot, which is also collected in an unstructured way.

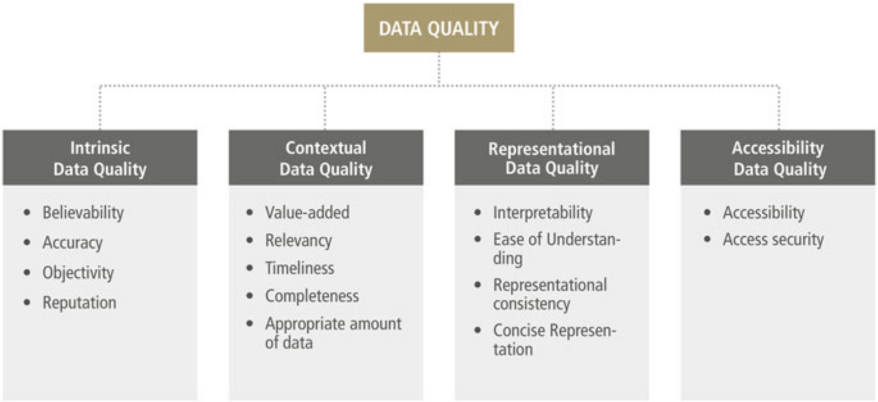

Fig. 1 Data quality dimensions according to Wang et al. (Source: Wang & Strong, 1996)

3 Data Quality Criteria

In order to optimize the value of corporate data, measures to improve data quality are of central importance in data management and data governance. However, despite numerous descriptions in the literature, there is no commonly accepted definition for the term “data quality”.[4] Olson explains: “data has quality if it satisfies the requirements of its intended use. It lacks quality to the extent that it does not satisfy the requirement. [...] To satisfy the intended use, the data must be accurate, timely, relevant, complete, understood, and trusted.” (Olson, 2003). Data quality is thus a metric that provides information about how well existing data can be applied in specific applications or business processes. That is, how well users of the data can apply the data for the respective applications’ purpose (“fitness for use”) (Wang & Strong, 1996). This fitness for use can change due to the varying needs of the users. Data quality thus depends on the time of observation and the available usage context (Weber & Klingenberg, 2020).

From a more technical perspective, data is of high quality if it is “error-free” and “meets specifications” (Fürber, 2016). Data quality is not a single characteristic, but multidimensional and context-dependent, i.e. “a set of data quality attributes that represent a single aspect or construct of data quality” (Wang & Strong, 1996, p. 6). Literature and practice provide a variety of different data quality dimensions (see, e.g., DAMA International, 2009; IQ International, 2017), with many approaches going back to Wang et al.’s dimensions published as early as 1992 (Wang et al., 1992). These were later condensed to 15 dimensions (Wang & Strong, 1996), which fall into four categories: intrinsic, contextual, representational, and accessibility (see Fig. 1).

[4] See, e.g., Ehrlinger et al. (2019), or Zulkiffli et al. (2019) for a discussion of terms.

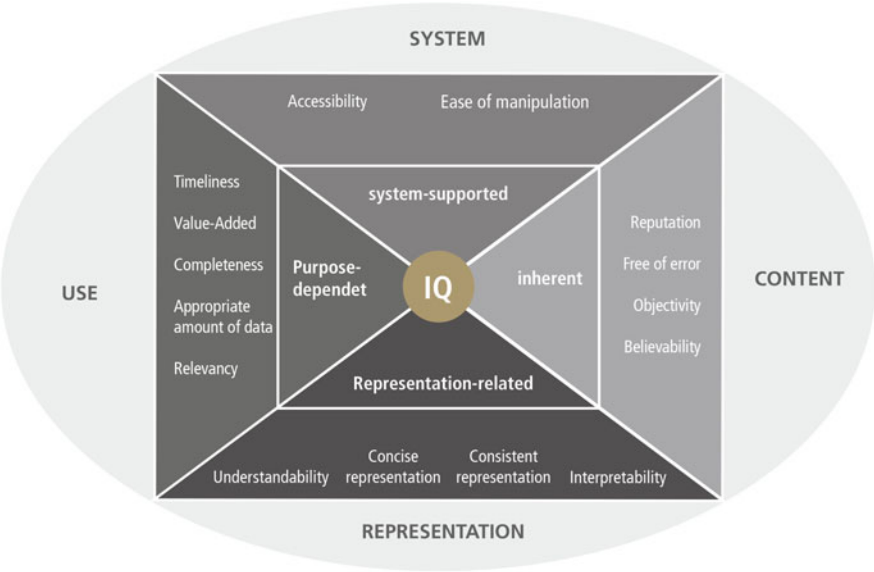

Fig. 2 15 dimensions of information quality according to DGIQ (Source: Rohweder et al., 2008)

Intrinsic data quality includes dimensions that deal with the quality of the data itself, such as accuracy or believability. Contextual data quality includes quality characteristics only observable when data is used. For example, relevancy can only be assessed in the context of the task at hand. The category of representational data quality includes dimensions related to the format and meaning of the data, such as representational consistency or ease of understanding. The category of accessibility considers dimensions related to the accessibility to data and access security.

Based on Wang et al. (1992) and a survey of IT users, the German Society for Information and Data Quality (DGIQ) developed a user-related definition (see Fig. 2), which categorizes data quality properties according to the (technical) system, the representation of the data, data use and content (Rohweder et al., 2008).

Since some of the definitions are fuzzy and, therefore, difficult to measure, dealing with overlapping criteria can be a costly problem. Consequently, practitioners often limit the assessment of data quality to a few selected dimensions—mostly focusing on the first four of the following dimensions:

- *Correctness/error-free:* Data is correct if it factually matches the attributes of the real-world reference object.
- *Completeness:* Data is complete if none of the necessary properties of the real-world object is missing.
- *Consistency:* Data is consistent if the attribute values of a data record are logically consistent and do not have logical inconsistencies with other data records of the same real-world object across different systems and sources.

- *Timeliness:* Data is up-to-date if it corresponds to the current state of the real-world object and represents its properties in a timely manner.
- *Availability/accessibility:* Data is available if it can be accessed by data users directly at the desired time.

4 From Content Accuracy to Mission-Critical Quality

Typically, data quality problems occur when requirements are not fulfilled. Data of poor quality implies duplicates, invalid or missing values. It is incomplete, erroneous, differently formatted, contradictory, and much more. Data quality problems are thus the direct result of unmet data requirements. Rahm and Do (2000) classify possible data errors by data source (see Fig. 3). The authors distinguish on a first level whether the problems occur within a single data source or during the integration of data from two or more sources. On a second level, whether problems affect the schema level or instance level.

The causes and reasons for poor data quality are manifold. In the age of social media and Big Data, the amount of data collected and the number of channels that can be used for collecting are constantly increasing. At the same time, data is often short-lived and changes quickly. Also, manual input often leads to incorrect or duplicate entries. For instance, when in a call center the name of the caller on the phone is misunderstood and thus re-entered in the system or not linked to existing customer information. In e-commerce, on average 19% of the customers do not remember their login data and therefore create a new account (Baymard Institute,

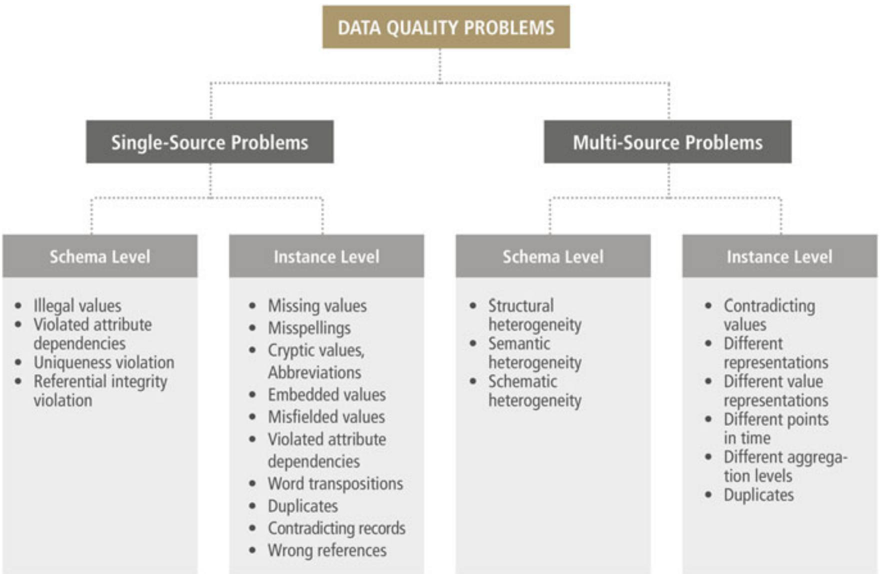

Fig. 3 Classification of data errors (Source: Rahm & Do, 2000)

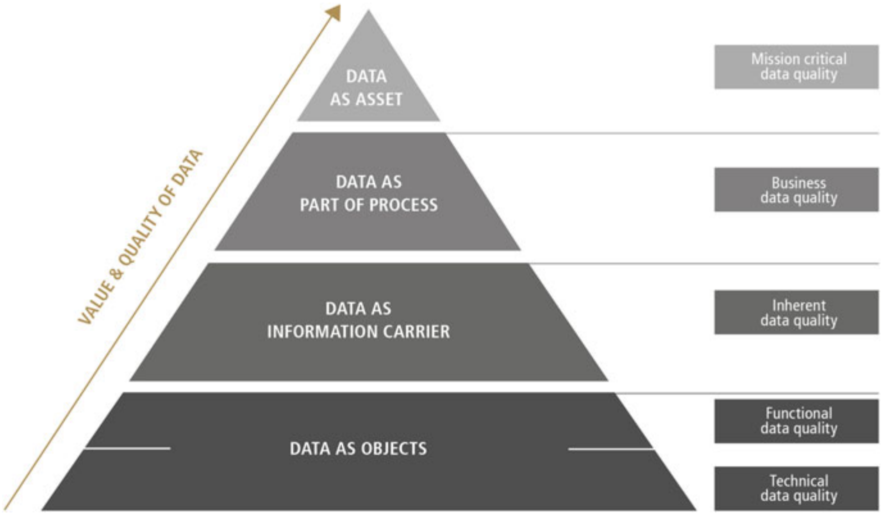

Fig. 4 Data quality and value pyramid (Source: Uniserv 2016a)

2017). When entering an address, users often select the country "Albania" because it appears at the top of the drop-down menu and can therefore be selected most quickly.

Within a company, a lack of communication between departments, too little awareness of data quality, difficulties in transferring data from other systems, a lack of technical support, or clear processes and guidelines for handling data can lead to poor data quality (Kraus, 2016). Rennhak (2006, p. 147) groups the different causes into the categories:

(a) Data entry—including incorrect, irrelevant, redundant, or even incomplete information
(b) Data transmission—technical errors or faulty data processing procedures during the transmission
(c) Data integration—flawed unification and consolidation processes
(d) Data storage—obsolete data attributes

Any of the above causes can hinder the success of a personalized cross-channel marketing strategy. When talking about data quality in relation to customer data, concepts such as address validation and duplicate cleansing, or matching come up quickly. But data quality is much more than just checking and, if necessary, correcting addresses or matching similar data records.

As the data quality and value pyramid illustrates, the quality of data can be ensured on four levels (see Fig. 4) and following up on all four will make data the ultimate asset for any company.

The first level is about technical and professional quality. Regarding technical quality, the data is viewed as a purely technical object without considering content.

Typical questions are, for example: Are data fields filled? Are the fields filled with the expected characters and in the correct format?

Functional quality is about content, but without looking at its relevance to business processes or applications. At this point, it must be ensured that the data makes sense, is admissible, and correct. For example, "February 31" would be technically correct but not meaningful and admissible as a date. A validation of dependencies must also be carried out at this stage, e.g., if the "Street" field is filled, the "House number" field must also be filled for German addresses. Otherwise, it cannot be considered a valid German address.

Validations at the first level can be automated, e.g., based on rules, and are ideally accompanied by error correction measures. These can be an automatic cleanup or, in ambiguous cases, a manual revision by data stewards. Another option would also be to route the data back to the person who created it so that they can make the correction.

At the second level, the focus is on validating that data is inherent. This includes address validation and duplicate cleansing. Only after it is ensured that a field for a telephone number is filled with correct characters in the specified format, a validation of the correctness of the telephone number is meaningful and necessary. Questions at this level can be: Is the address correct? Is the specified e-mail address accurate? Are there duplicates and how can they be merged?

Only inherent data can be a reliable basis for business processes. For example, article numbers are not just a piece of information in an order but must also be assigned to the right product. Only then, the order processing or shipping preparation can start.

To that end, rule-based validations of the content against specified references may be useful, which demand in-depth knowledge of business processes and business logic. Close collaboration with or direct involvement of the departments that have the required business know-how is indispensable for the successful development of such rules.

The eventual validation process for substantial correctness can largely be automated. Special validation processes, such as address validation or checking for duplicates, can be carried out in a similar way. For example, data about prospects transmitted from an internet portal can automatically be validated against existing data and, if necessary, marked as potential duplicates. Wrong data records that are neither automatically nor manually cleansed must be ejected so that they do not lead to errors, additional work, or inappropriate decisions in subsequent processes.

Once it is ensured that the data meets the basic quality requirements, the third level can address the question of their business utility. This step depends on the individual requirements of a company. Examples include finding out which products are in high demand in certain regions or which orders could not be fulfilled due to a lack of availability and what sales losses were associated with this. Are insights into cross-selling and up-selling potentials of interest? Legal requirements, such as age controls and embargoes, can also be monitored in this way. If the existing data cannot answer a company's questions, this may be due to incorrect or missing data.

In this case, the data on the prior levels must be revised in accordance with the requirements.

Finally, the fourth level contains data that supports mission-critical decisions. These are, for example, investments made based on predictive analytics. For this purpose, the data must meet the highest quality standards in order to minimize the risk of wrong decisions. If data meets these requirements, one can speak of data as an asset that creates added value for the company. Further steps on the way to an "all-encompassing data truth" in the context of customer data are, first, the "Golden Record" and, second, the so-called Golden Profile (Uniserv 2016b).

The Golden Record consolidates the existing information on a person, such as the postal address, e-mail addresses, telephone numbers, social login, and bank details, from various sources into a master data record. The Golden Profile enriches this record with transaction and interaction data. The Golden Profile is, therefore, composed of the following data:

- Customer behavior (e.g., orders, payment history, dwell time)
- Data describing the person (e.g., attributes, self-disclosures, demographics)
- Customer characteristics (e.g., opinions, preferences, needs, wants)
- Customer interactions (e.g., offers, results, context, clickstreams, notes)

With the Golden Profile, a complete, accurate, and up-to-date 360° view of customers is possible, enabling companies to make offers adapted to individual needs, as they are expected by customers in the digital age.

The Golden Profile can be viewed and further developed as the customer's digital twin (Braun et al., 2022). For example, detailed transaction data can be combined with interaction data from social media and usage or consumption data from Internet-of-Things-enabled products and devices to create a 360° view of customers that is expanded by many additional attributes (Hechler et al., 2020). This creates new opportunities to model customers more accurately in order to adapt products and services and improve the customer experience.

5 Data Quality Assessment

Ensuring a high quality of data so that it becomes a valuable asset is not a one-time process. As data changes over time, so does its quality. Making it a necessity that data quality is re-evaluated again and again. Deutsche Post Direkt GmbH (2021), for example, found in its "Address Study 2021" study that around 14 million relocations, 990,000 deaths, as well as 370,000 marriages and 150,000 divorces resulted in a large number of address and name changes every year. These changes, as well as changes in street names or changes in company names due to mergers and acquisitions, have to be maintained.

A company must know which data is collected, stored, and used by which systems for which purposes. In other words, data quality must be secured throughout the entire data life cycle. This requires specific organizational structures. This is

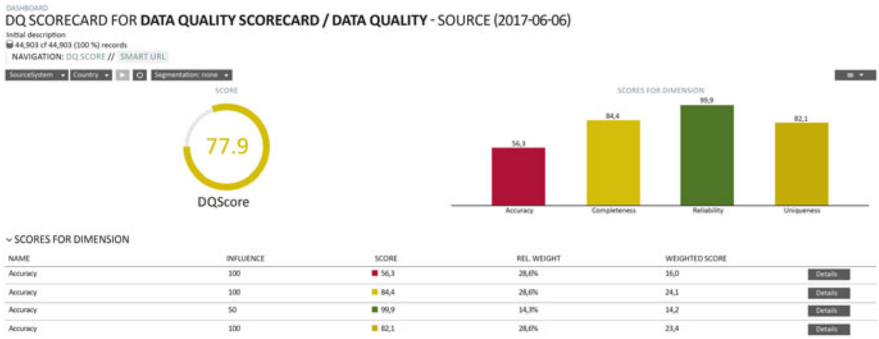

Fig. 5 Data Quality Scorecard of Uniserv GmbH

where data governance comes into play. It defines responsibilities, processes, standards, and metrics that ensure the efficient use of data as an economic asset and thus support the achievement of the company's strategic goals (Weber, 2012; DEMAND, 2019).

Data governance is a shared responsibility of corporate management, strategy, and specialist departments, such as customer service, sales, marketing, and IT. During its implementation, it is necessary to consolidate the knowledge of business processes and application systems distributed over the various corporate divisions. This goes hand in hand with a centralized definition and maintenance of business rules and a continuous monitoring of these rules during operations. In addition, companies need to keep a strategic and an operational perspective on data quality.

To ensure that the data precisely meets the required qualitative specifications, individual and application-specific business rules are defined that describe characteristics of data within the specific corporate contexts. These rules can then be used to validate entire data sets or individual data records. The rules thus operationalize the requirements that data users have on the data (Klingenberg & Weber, 2017).

Rules can address either one or more attributes of a data record. However, they can also refer to several data records across attributes and thus encompass the previously mentioned different levels of data quality. Rules are typically recorded for questions such as: "Are there empty data fields?" "Do the individual attributes, such as street, house number, postal code, and city, result in an accurate and existing address?" or: "Does each customer exist only once in the system?" If you want to check whether the existing customer data is sufficient for an e-mail marketing campaign, you need rules that evaluate whether the e-mail address field is maintained regularly.

For evaluating compliance with these rules, corresponding metrics are defined. These must be measured and monitored regularly. This is possible, for example, with a data quality (DQ) scorecard tool, as offered by various providers such as Uniserv GmbH (see Fig. 5). With a DQ scorecard, the entire data set can be validated against individual business rules. The validation results can be weighted and analyzed according to individual rules and are aggregated on different levels for a better overview and understanding. For example, custom metrics can be generated with the Uniserv DQ Scorecard by aggregation across a group of fields such as name and address elements to determine if these records are unique. The result is the Data Quality Score, which assesses the overall quality of all customer records across all defined rules.

During validation a distinction is made between rules that are "fulfilled," "not fulfilled," and "not applicable." For example, a syntax validation of an e-mail address would be "not applicable" if a previous rule shows that the field is not filled. Having such a "not applicable" status prevents multiple negative validations of a specific data attribute.

In addition, rules can be combined into rule groups and aggregated into hierarchy levels for these, for which a single data quality score can be calculated. For example, all rules related to address data are assigned to an "Address" group and all rules related to personal customer data are assigned to the "Personal data" group.

Drill-down functionality at group and field (i.e., attribute) level allows to see exactly which rules or entities have led to a lower data quality score—in other words, it shows where problems and weak points are hidden. Based on this knowledge, companies can define targeted measures to optimize data quality. By comparing different data quality scores, it is possible to evaluate which measures have led to an improvement in data quality. A comparison is also possible across different data sources, applications, and business areas.

6 Ensure Data Quality with the Help of the DQ Scorecard: A Practical Example

The DQ Scorecard allows data stewards to evaluate data quality in an easy and straightforward way, as presented in the following example.

A company sells household products worldwide. Direct access to customers is of particular importance. The products are sold via a large network of independent consultants. In addition, selected products are offered online and in stores. However, regardless of the sales channel, the product consultant remains the customer's first point of contact.

The subsidiaries in the individual countries operate largely independently. As a result, a complex system and process landscape that provides customer data has evolved over the years. Each country subsidiary utilizes its own databases and applications adapted to country-specific needs. In such a heterogeneous environment, professional customer data management is crucial to success.

For example, marketing campaigns must be optimally tailored to the needs of the customer base. A 360° customer view is recommended for effective customer loyalty measures. In order to create a golden customer profile, clarity about possible data quality deficiencies must be achieved. In our scenario, a heterogeneous system landscape might additionally complicate a cross-national overview of customer master data quality.

The DQ Scorecard from Uniserv is a solution that determines the quality of the customer master data for marketing measures across a complex system and process landscape. The first step is to create a uniform global view of the data from the various countries. This involves defining a uniform data structure with predefined fields so that the data landscape can be transformed into one consistent data format. This is followed by the definition of specific business rules that are used to validate the data records. This rule-based validation process forms the main component of the solution. The result is the so-called Data Quality Score. The Data Quality Score makes symptoms of bad data visible and measurable. Part of the definition is also how intermediate results are combined into the overall score, and with what weighting. Finally, the resulting set of rules is applied to the "global view" of all customer master data of all countries. The outcome of the rule check can be tracked on a browser-based dashboard. This gives decision-makers at the company the opportunity to efficiently compare and analyze the individual data quality scores.

In this way, the marketing department gains an overview of the current, over-arching status quo of customer data quality across all subsidiaries and countries. Targeted measures to improve quality can be initiated, thus creating the prerequisite for the formation of Golden Profiles. With the help of the DQ Scorecard, employees in each country-specific organization can identify optimization potential, monitor the effectiveness of data quality measures, and initiate further improvement measures. Awareness about data and its quality can thus be significantly increased and, in turn, the data quality score permanently optimized. Marketing campaigns can be planned more effectively, as it is clearly visible whether the customer master data meets the respective requirements of the marketing campaign or not.

7 Summary

Together with digital transformation also customer data is moving into the focus of companies. As a result, data is becoming the number one competitive factor. Establishing good data practices requires not only long-term decisions by upper management but also daily decisions at the middle and lower levels of a company. The four-level pyramid shows how important adequate data quality is to utilize data as a valuable asset. If the data quality is inappropriate, corporate management decisions will be based on the results of incorrect data analysis and the digital transformation will be doomed to failure.

Akshay Tandon, Vice President and Head of Strategy and Analytics at LendingTree Inc., an American online financial services provider, also confirms the central importance of high-quality customer data in his statement (Burns, 2017): "It's like when you see a skyscraper: You're impressed by the height, but nobody is impressed by the foundation. But make no mistake, it's important. You have to have good data management to take advantage of AI."

Consequently, successful planning of future marketing and sales campaigns with the help of machine learning and predictive analytics is only possible if a 360° view of customers can be obtained. The "Golden Profile" integrates previously distributed customer data into such a 360° customer view and turns data into a valuable asset for any company.

Since the quality of the data can change over time, it should be measured continuously. A data quality scorecard clearly shows the status of data quality and where potential for optimization is hidden. Finally, a data quality score reveals the actual value of data for companies.

References

Azkan, C., et al. (2019). *Readiness Data Economy. Bereitschaft der deutschen Unternehmen für die Teilhabe an der Datenwirtschaft*. IW-Gutachten.

Baymard Institute. (2017). *Checkout optimization and reducing abandonments: 6 Ways to retain shoppers from 'add to cart' to order completion*. https://pages.amazonpayments.com/rs/846-RQB-314/images/Baymard%20Report%20-%20US.pdf

Bernhard, M., & Mühling, T. (2020). *Verantwortungsvolle KI im E-Commerce: Eine kurze Einführung in Verfahren der Künstlichen Intelligenz in der Webshop-Personalisierung*. Springer-Gabler.

Berthold, M. R., Borgelt, C., Höppner, F., Klawonn, F., & Rosaria, S. (2020). Data preparation. In *Guide to intelligent data science. Texts in computer science* (pp. 127–156). Springer.

Bognar, Z., Puljic, N., & Kadezabek, D. (2019). *Impact of influencer marketing on consumer behaviour*. Proceedings of the 42nd International Scientific Conference on Economic and Social Development, London, June 12–13, 2019, pp. 301–309. Varazdin Development and Entrepreneurship Agency (VADEA).

Bitkom. (2020, Juli). *E-Commerce und stationärer Handel: So digital shoppen die Deutschen. Ein Bitkom-Studienbericht*.

BMWi. (2017). *Weißbuch Digitale Plattformen.* Accessed March 20, 2017, from https://www.bmwi.de/Redaktion/DE/Publikationen/Digitale-Welt/weissbuch-digitale-plattformen.html

Braun, S., Follwarczny, D., & Heißler, A. (2022). Neue Kanäle – neue Daten: Die veränderte Rolle von Kundendaten im Handel. In T. Breyer-Mayländer, C. Zerres, A. Müller, & K. Rahnenführer (Eds.), *Die Corona-Transformation – Krisenmanagement und Zukunftsperspektiven in Wirtschaft, Kultur und Bildung* (pp. 133–159). Springer-Gabler.

Bult, J. R., & Wansbeek, T. J. (1995). Optimal selection for direct mail. *Marketing Science, 14*(4), 378–394.

BVDW – Bundesverband Digitale Wirtschaft. (2018). *Data economy*. Datenwertschöpfung und Qualität von Daten.

Burns, E. (2017). *Search Business Analytics: Analytics VP: AI projects must be built on a good data foundation*. http://searchbusinessanalytics.techtarget.com/feature/Analytics-VP-AI-projects-mustbe-built-on-a-good-data-foundation

DAMA International. (2009). *The DAMA Guide to the Data Management Body of Knowledge (DAMADMBOKGuide)*. Technics Publications.

Data IQ. (2017). *Royal Mail: 6 % Solution = How better customer data drives Marketing performance and business growth*. https://www.royalmail.com/sites/default/files/RMDS-Research-Report-2017-Six-per-cent-solution.pdf

DEMAND. (2019). *Data Economy. Status quo der deutschen Wirtschaft & Handlungsfelder in der Data Economy*. https://www.demand-projekt.de/

Deutsche Post Direkt. (2021). *Adress-Studie 2021. Untersuchung zur Qualität von Kundenadressen in Deutschland*. https://www.deutschepost.de/content/dam/dpag/images/D_d/DDP/Downloads/studien/dp-adress-studie-2021.pdf
Eichsteller, H., & Seitz, J. (2019). *Studie Digital Dialog Insights 2019*. Bundesanzeiger Verlag.
Ehrlinger, L., Rusz, E., & Wöß, W. (2019). *A survey of data quality measurement and monitoring tools*. CoRR abs/1907.08138, https://arxiv.org/abs/1907.08138
European Union. (2020). *The European data market monitoring tool: Key facts & figures, first policy conclusions, data landscape and quantified stories*. D2.9 Final Study Report. Publications Office of the European Union.
Fürber, C. (2016). *Data quality management with semantic technologies*. Springer.
Gafni, R., & Golan, O. T. (2016). The influence of negative consumer reviews in social networks. *Online Journal of Applied Knowledge Management, 4*(2), 44–58.
Gentsch, P. (2017). *Künstliche Intelligenz für Sales, Marketing und Service. Mit AI und Bots zu einem Algorithmic Business – Konzepte, Technologien und Best Practices*. Springer Fachmedien.
Google. (2016). *The B2B search landscape: Targeting the stages of your buyer journey to drive revenue*. Big Footprint Google Event, Oktober 2016. https://www.bigfootprintdigital.com/wpcontent/uploads/2016/10/Big-Footprint-Google-Event-FINAL-DECK.pdf
Greven Medien & GfK. (2019). *GfK-Online-Umfrage*. https://www.greven.de/bewertungen-gfk
Hechler, E., Oberhofer, M., & Schaeck, T. (2020). Applying AI to master data management. In *Deploying AI in the enterprise*. Apress. https://doi.org/10.1007/978-1-4842-6206-1_9
ISO International Organization for Standardization. (2016). *ISO 8000-100:2016 Data quality – Part 100: Master data: Exchange of characteristic data: Overview*. Standard. https://www.iso.org/standard/62392.html
IQ International. (2017). *Fundamentals of IQ*. International Association for Information and Data Quality. https://www.iqint.org/knowledge/fundamentals/
Kessler, R., & Gómez, J. M. (2020). Implikationen von Machine Learning auf das Datenmanagement in Unternehmen. *HMD Praxis der Wirtschaftsinformatik, 57*, 89–105.
Klingenberg, C., & Weber, K. (2017). Kundendatenqualität im Zeitalter der digitalen Transformation. In *IT-Governance. Fachzeitschrift des ISACA Germany Chapter e.V.* (Vol. 26, pp. 7–13).
Kraus, C. (2016). *Warum ein Online-Shop gute Datenqualität braucht*. COMPUTERWOCHE. Accessed January 02, 2016, from https://www.computerwoche.de/a/warum-ein-online-shop-gute-datenqualitaet-braucht,3217296
Kreutzer, R., & Land, K.-H. (2013). *Digitaler Darwinismus*. Springer.
Kröger, J., & Marx, S. (2020). *Marketing im Wandel. In Agile Marketing* (pp. 63–81). Springer Gabler.
Lünendonk, J. (2019). *Lünendonk-Studie 2019 – Künstliche Intelligenz.* Lünendonk & Hossenfelder GmbH in Zusammenarbeit mit GINKO Consulting Management.
Maycock, M., Spenner, P., Hamshar, R., & West, S. (2012). *Marketing leadership council: The digital evolution in B2B marketing*. http://www.thebluecollarmarketer.com/wp-content/uploads/2015/10/the-digital-evolution-in-b2b-marketing_research-studies.pdf
Myler, L. (2017). *Forbes: Better data quality equals higher marketing ROI*. Accessed July 12, 2017, from https://www.forbes.com/sites/larrymyler/2017/07/12/better-data-quality-equals-higher-marketingroi/#48e520b7b68d
Olson, J. (2003). *Data quality: The accuracy dimension*. Morgan Kaufmann; Elsevier Science.
Otto, B., & Österle, H. (2016). *Corporate Data Quality. Voraussetzung erfolgreicher Geschäftsmodelle*. Springer-Gabler.
Piro, A., & Gebauer, M. (2021). Definition von Datenarten zur konsistenten Kommunikation im Unternehmen. In K. Hildebrand, M. Gebauer, & M. Mielke (Eds.), *Daten- und Informationsqualität*. Springer Vieweg. https://doi.org/10.1007/978-3-658-30991-6_8
Rahm, E., & Do, H.-H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin, 23*(4), 3–13.

Rennhak, C. (2006). *Herausforderung Kundenbindung*. Gabler Edition Wissenschaft. Springer-Verlag.
Rohweder, J., Kasten, G., Malzahn, D., Piro, A., & Schmid, J. (2008). Informationsqualität – Definitionen, Dimensionen und Begriffe. In K. Hildebrand, M. Gebauer, H. Hinrichs, & M. Mielke (Eds.), *Daten- und Informationsqualität*. Vieweg+Teubner Verlag.
SPLENDID RESEARCH. (2018). *Influencer: Einfluss auf die Kaufwahrscheinlichkeit*. Hamburg, Juli 2019. https://www.splendid-research.com/studie-influencer.html
Stern, P., Büllesfeld, E., & Bierkandt, J. (2021, March). *KI KI im Kundendatenmanagement – Einblicke in Anwendungsmöglichkeiten*. http://publica.fraunhofer.de/dokumente/N-630689.html
Stevens, J. (2019). Datenqualität bei algorithmischen Entscheidungen. In K. David, K. Geihs, M. Lange, & G. Stumme (Eds.), *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik* (pp. 367–380). Informatik für Gesellschaft.
Uniserv GmbH. (2016a). *Data Quality ist mehr als Adressvalidierung*. White Paper.
Uniserv GmbH. (2016b). *Digitale Transformation needs Kundendatenmanagement*. White Paper.
Uniserv GmbH. (2019). *Trendstudie Customer Data Management 2019*. Accessed June 06, 2019, from https://www.uniserv.com/unternehmen/blog/detail/article/ergebnisse-der-trendstudie-kundendatenmanagement-2019/
Uniserv GmbH. (2020). *Praxisstudie Customer Data Management 2020*. Accessed July 29, 2020, from https://www.uniserv.com/unternehmen/blog/detail/article/management-von-geschaeftspartnerdaten-nicht-oberste-prioritaet-bei-unternehmen/
van Bommel, E., Edelman, D., & Ungerman, K. (2014). *Digitizing the consumer decision journey*. McKinsey. http://www.mckinsey.com/business-functions/marketingandsales/our-insights/digitizing-theconsumer-decision-journey
Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.
Wang, R., Reddy, M., & Kon, H. (1992). *Toward quality data: An attribute-based approach*. Massachusetts Institute of Technology (MIT). http://web.mit.edu/smadnick/www/wp2/1993-04-SWP%233515.pdf
Weber, K. (2012). Data Governance – Organisation des Stammdatenmanagements. *IT-Governance, 6*(13), 3–8.
Weber, K., & Klingenberg, C. (2020). *Data Governance*. Carl Hanser Verlag GmbH & Co. KG.
Weber, K., Otto, B., & Lis, D. (2021). Data governance. In K. Hildebrand, M. Gebauer, & M. Mielke (Eds.), *Daten- und Informationsqualität* (pp. 271–291). Springer-Gabler.
Wolters, H. M. (2020). Qualität von Kundendaten – Ansätze zur Analyse und Verbesserung als Basis für effiziente Marketingentscheidungen. In S. Boßow-Thies, C. Hofmann-Stölting, & H. Jochims (Eds.), *Data-driven marketing* (pp. 29–41). Springer-Gabler.
Zulkiffli, P. N. I. N., Akhir, E. A. P., Aziz, N., & Cox, K. (2019). The development of data quality metrics using thematic analysis. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*, 304–310.

Simone Braun is Professor of E-Commerce at Offenburg University of Applied Sciences. She teaches and researches in the field of omni-channel commerce, customer data management, and AI in retail and e-commerce. She has many years of experience in the field of innovation development with a special focus on customer data.

Andreas Heißler has worked for Uniserv GmbH, Pforzheim, for more than 20 years. His tasks in communications include press and public relations, content, lead management, and data protection. He writes about the importance of data as a valuable asset for companies in various industries, in the context of digital transformation and artificial intelligence.