Requirements for Data Quality Metrics

BERND HEINRICH, University of Regensburg
DIANA HRISTOVA, University of Regensburg
MATHIAS KLIER, University of Ulm
ALEXANDER SCHILLER, University of Regensburg
MICHAEL SZUBARTOWICZ, University of Regensburg

Data quality and especially the assessment of data quality have been intensively discussed in research and practice alike. To support an economically oriented management of data quality and decision-making under uncertainty, it is essential to assess the data quality level by means of well-founded metrics. However, if not adequately defined, these metrics can lead to wrong decisions and economic losses. Therefore, based on a decision-oriented framework, we present a set of five requirements for data quality metrics. These requirements are relevant for a metric that aims to support an economically oriented management of data quality and decision-making under uncertainty. We further demonstrate the applicability and efficacy of these requirements by evaluating five data quality metrics for different data quality dimensions. Moreover, we discuss practical implications when applying the presented requirements.

CCS Concepts: • **Information systems~Data management systems.**

Additional Key Words and Phrases: Data quality, Data quality assessment, Data quality metrics, Requirements for metrics

1 INTRODUCTION

Due to the rapid technological development, companies increasingly rely on data to support decision-making and to gain competitive advantage. To make informed and effective decisions, it is crucial to assess and assure the quality of the underlying data. 83% of the respondents of a survey conducted by Experian Information Solutions [2016] state that poor data quality has actually hurt their business objectives, and 66% report that poor data quality has had a negative impact on their organization in the last twelve months. Another report reveals that 84% of the CEOs are concerned about the quality of the data they use for decision-making [KPMG 2016; Forbes Insights 2017]. In addition, Gartner indicates that the average financial impact of poor data quality amounts to $9.7 million per year and organization [Moore 2017]. Overall, it is estimated that poor data quality costs the US economy $3.1 trillion per year [IBM Big Data and Analytics Hub 2016]. In the light of the current proliferation of big data with large amounts of heterogeneous, quickly-changing data from distributed sources being analyzed to support decision-making, assessing and assuring data quality becomes even more relevant [IBM Global Business Services 2012; Buhl et al. 2013; Cai and Zhu 2015; Flood et al. 2016]. Indeed, the three characteristics Volume, Velocity and Variety, often called the three Vs of big data, make the assurance of data quality increasingly challenging (e.g., due to the integration of various data sources or when considering linked data; cf. also Cappiello et al. 2016; Debattista et al. 2016). Thus, the consequences of wrong decisions are becoming even more costly [SAS Institute 2013; Forbes Insights 2017]. This has resulted in the addition of a fourth V (=Veracity) reflecting the importance of data quality in the context of big data [Lukoianova and Rubin 2014; Flood et al. 2016; IBM Big Data and Analytics Hub 2016].

Data quality can be defined as “the measure of the agreement between the data views presented by an information system and that same data in the real world” [Orr 1998, p. 67; cf. also Parssian et al. 2004; Heinrich et al. 2009]. Data quality is a multi-dimensional construct [Redman 1996; Lee et al. 2002; Eppler 2003; Taleb et al. 2016] comprising different data quality dimensions such as accuracy, completeness, consistency and currency [Wang et al. 1995; Batini and Scannapieco 2016]. Each data quality dimension provides a particular perspective on the quality of data views. As a result, researchers have developed corresponding metrics for the quantitative assessment of these dimensions for data views [e.g., Ballou et al. 1998; Hinrichs 2002; Even and Shankaranarayanan 2007; Heinrich et al. 2007; Fisher et al. 2009; Heinrich et al. 2009; Blake and Mangiameli 2011; Heinrich et al. 2012; Wechsler and Even 2012; Heinrich and Klier 2015; Heinrich and Hristova 2016]. Metrics assessing such data quality dimensions for data views and data values stored in IS are in the focus of this paper. In contrast, for instance metrics addressing the quality of data schemes are not directly considered.

Data quality metrics provide measurements for data views with greater (lower) metric values representing a greater (lower) level of data quality and each data quality level being represented by a unique metric value. They are needed for two main reasons. First, the metric values are used to support data-based decision-making under uncertainty. Here, well-founded data quality metrics are required to indicate to what extent decision makers should rely on the underlying data values. Second, the metric values are used to support an economically oriented management of data quality [cf., e.g., Wang 1998; Heinrich et al. 2009]. In this context, data quality improvement measures should be applied if and only if the benefits (due to higher data quality) outweigh the associated costs. To be able to analyze which data quality improvement measures are efficient from an economic perspective, well-founded data quality metrics are needed to assess (the changes in) the data quality level.

While both research and practice have realized the high relevance of well-founded data quality metrics, many data quality metrics still lack an appropriate methodical foundation as they are either developed on an ad hoc basis to solve specific problems [Pipino et al. 2002] or are highly subjective [Cappiello and Comuzzi 2009]. Hinrichs [2002], for example, defines a metric to assess the correctness of a stored data value ω as $DQ(\omega,\omega_m) := \frac{1}{d(\omega,\omega_m)+1}$ where ω_m represents the the corresponding real-world value and d a domain-specific distance measure. For instance, as proposed by Hinrichs [2002], let $d(\omega,\omega_m)$ be the Hamming distance between the stored and the correct value (i.e., the number of positions at which the corresponding symbols of two data strings are different). Applying this metric to $(\omega,\omega_m) = (\text{‘Jefersonn’},\text{‘Jefferson’})$ and $(\omega,\omega_m) = (\text{‘Jones’},\text{‘Adams’})$ to determine the correctness of customers' surnames in a product campaign yields the following results: $DQ(\text{‘Jefersonn’},\text{‘Jefferson’}) = \frac{1}{5+1} \approx 16.67\%$ and $DQ(\text{‘Jones’},\text{‘Adams’}) = \frac{1}{4+1} = 20\%$. If the decision criterion in the product campaign is a metric value of at least 20%, a sales letter is sent to ‘Jones’, which will most probably not reach its destination, whereas no sales letter is sent to ‘Jefersonn’, which would much more likely reach its destination. To avoid such problems, both researchers and practitioners set out to propose requirements for data quality metrics [e.g., Pipino et al. 2002; Even and Shankaranarayanan 2007; Heinrich et al. 2007; Mosley et al. 2009; Loshin 2010; Hüner 2011]. Most of them, however, did not aim at justifying the requirements based on a decision-oriented framework. As a result, the literature on this topic is fragmented and it is not clear which requirements are indeed relevant to support decision-making. Moreover, as some of the requirements leave room for interpretation, their verification is difficult and subjective. This results in a research gap which we aim to address by answering the following research question:

> Which clearly defined requirements should a data quality metric satisfy to support both decision-making under uncertainty and an economically oriented management of data quality?

To address this research question, we propose a set of five requirements, namely the existence of minimum and maximum metric values (R1), the interval scaling of the metric values (R2), the quality of the configuration parameters and the determination of the metric values (R3), the sound aggregation of the metric values (R4), and the economic efficiency of the metric (R5).

We analyze existing literature and justify this set of requirements based on a decision-oriented framework. As a result, our requirements support both decision-making under uncertainty and an economically oriented management of data quality. Data quality metrics which do not meet them can lead to wrong decisions and/or economic losses (e.g., because the efficiency of the metric's application is not ensured). Moreover, the presented requirements facilitate a well-founded assessment of data quality, which is crucial for supporting data governance initiatives [Weber et al. 2009; Khatri and Brown 2010; Otto 2011; Allen and Cervo 2015] and an efficient data quality management [cf. also Cappiello and Comuzzi 2009; Fan 2015].

The need for such requirements is further supported by the discussions in other fields of research such as software engineering. For example, Briand et al. [1996] provide a universal set of properties for the sound definition of software measures. The proposed properties can be used by researchers to “validate their new measures” (p. 2) and can be interpreted as necessary requirements for software metrics. In addition, in the context of ISO/IEC standards the SQuaRE series aims to “assist those developing and acquiring software products with the specification and evaluation of quality requirements” [p. V in ISO/IEC 25020 2007; cf. also Azuma 2001]. In particular, ISO/IEC 25020 provides criteria for selecting software quality measures with the same motivation as above.

The remainder of the paper is structured as follows. In the next section, we provide an overview of the related work and identify the research gap. Section 3 comprises the decision-oriented framework for our work. In Section 4, we propose a set of five requirements for data quality metrics which are defined and justified based on this framework. In Section 5, we demonstrate the applicability and efficacy of these requirements using five data quality metrics from literature. Section 6 contains a discussion of practical implications. The last section provides conclusions, limitations and directions for future research.

2 RELATED WORK

In this section, we analyze existing works, which propose requirements for data quality metrics. Following the guidelines of standard approaches to prepare the related work [e.g., Webster and Watson 2002; Levy and Ellis 2006], we searched the databases ScienceDirect, ACM Digital Library, EBSCO Host, IEEE Xplore, and the AIS Library as well as the Proceedings of the International Conference on Information Quality (ICIQ) for the following search term and without posing a restriction on the time period: *(“data quality” and metric* and requirement*) or (“data quality” and metric* and standard*) or (“information quality” and metric* and requirement*) or (“information quality” and metric* and standard*)*. This search led to 136 papers which were manually screened based on title, abstract, and keywords. The remaining 43 papers were analyzed in detail and could be divided into three disjoint categories A, B and C. Category A comprises requirements for data quality metrics and data quality metric values from a *methodical* perspective. Category B contains requirements concerning the *general data quality assessment process* in an *organization* (e.g., measurement frequency). Category C consists of requirements and (practical) recommendations for the *concrete organizational integration* of data quality metrics (e.g., within business processes). Regarding our research question, we focused on Category A comprising five relevant papers on which we performed an additional forward and backward search, resulting in a total of eight relevant papers discussed in the following.

Pipino et al. [2002] propose the functional forms *simple ratio*, *min or max operation*, and *weighted average* to develop data quality metrics. *Simple ratio* measures the ratio of the number of desired outcomes (e.g., number of accurate data units) to the total number of outcomes (e.g., total number of data units). *Min or max operation* can be used to define data quality metrics requiring the aggregation of multiple assessments, for instance on the level of data values, tuples, or relations. Here, the minimum (or maximum) value among the normalized values of the single assessments is calculated. *Weighted average* is an alternative to the *min or max operation* and represents the weighted average of the single assessments. The major goal of Pipino et al. [2002] is to present feasible and useful functional forms which can be seen as a first important step towards requirements for data quality metrics. They ensure the range [0; 1] for the metric values and address the aggregation of multiple assessments.

Even and Shankaranarayanan [2007] aim at an economically oriented management of data quality. They propose four consistency principles for data quality metrics. *Interpretation consistency* states that the metric values on different data view levels (data values, tuples, relations, and the whole database) must have a consistent semantic interpretation. *Representation consistency* requires that the metric values are interpretable for business users (typically on the range [0; 1] with respect to the utility resulting from the assessed data). *Aggregation consistency* states that the assessment of data quality on a higher data view level has to result from the aggregation of the assessments on the respective lower level. The aggregated result should take values, which are not higher than the highest or lower than the lowest metric value on the respective lower level. *Impartial-contextual consistency* means that data quality metric values should reflect whether the assessment is context-dependent or context-free.

Heinrich et al. [2007; 2009; 2012] analyze how data quality can be assessed by means of metrics in a goal-oriented and economic manner. To evaluate data quality metrics, they define six requirements. *Normalization* requires that the metric values fall into a bounded range (e.g., [0; 1]). *Interval scale* states that the difference between any two metric values can be determined and is meaningful. *Interpretability* means that the metric values have to be interpretable, while *aggregation* states that it must be possible to aggregate metric values on different data view levels. *Adaptivity* requires that it is possible to adapt the metric to the context of a particular application. *Feasibility* claims that the parameters of a metric have to be determinable and that this determination must not be too cost-intensive. Moreover, this requirement states that it should be possible to calculate the metric values in an automated way.

Mosley et al. [2009] and Loshin [2010] discuss requirements for data quality metrics from a practitioners' point of view. Both contributions comprise the requirements *measurability* and *business relevance* claiming that data quality metrics have to take values in a discrete range and that these values need to be connected to the company's performance. Loshin [2010] adds that it is important to clearly define the metric's goal and to provide a value range and an interpretation of the parts of this range (*clarity of definition*). In addition, Mosley et al. [2009] require *acceptability*, which implies that a metric is assigned a threshold at which the data quality level meets business expectations. If the metric value is below this threshold, it has to be clear who is accountable and in charge to take improvement actions. The corresponding requirements *accountability/stewardship* and *controllability*, however, refer to the integration of a data quality metric within organizations (cf. Category C) and are thus not within the focus of this paper. The same holds for the requirements *representation* and *reportability* as found in both works and also *drill-down capability* by Loshin [2010]. *Representation* claims that the metric values should be associated with a visual representation, *reportability* points out that they should provide enough information to be included in aggregated management reports, and *drill-down capability* states that it should be possible to identify a data quality metric's impact factors within the organization. Finally, *trackability* which requires a metric to be repeatedly applicable at several points of time in an organization (cf. Category B) is also beyond the focus of this paper.

Hüner [2011] proposes a method for the specification of business-oriented data quality metrics to support both the identification of business critical data defects and the repeated assessment of data quality. Based on a survey among experts, he specifies 21 requirements for data quality assessment methods (cf. Appendix B). However, only some of them constitute methodical requirements for data quality metrics and metric values (cf. Category A) and are thus considered further. These are *cost/benefit*, *definition of scale*, *validity range*, *comparability*, and *comprehensibility*. The other requirements refer to Category B (e.g., *repeatability*, *definition of measurement frequency*, *definition of measurement point, definition of measurement procedure*) or Category C (e.g., *responsibility*, *escalation process*, *use in SLAs*) and are not within the focus of this paper.

To sum up, prior works provide valuable contributions by stating a number of possible requirements for data quality metrics and their respective values. While some of them overlap, existing literature is still very fragmented. In addition, many requirements are not clearly defined, which makes their application and verification very difficult. To address these issues, we organize the existing requirements in six groups with each group being characterized by a clear, unique characteristic (cf. Table 1). Note that some of the requirements which leave room for interpretation (cf. brackets in Table 1) are classified in more than one group. Further, some of these existing requirements (e.g., *simple ratio*, *weighted average*) could also be understood as a way to define a data quality metric. In the following, however, they are considered as requirements for data quality metrics. For example, *simple ratio* in Group 1 means that a data quality metric should attain values in [0; 1].

Table 1. Groups of Requirements

Group	Keyword	Requirements
1	range	*normalization, validity range, clarity of definition (range), simple ratio (bounded in [0; 1]), representation consistency (range), measurability*
2	scale	*interval scale, definition of scale (scale)*
3	interpretation	*interpretability, clarity of definition (interpretation), simple ratio (interpretation), interpretation consistency (interpretation), comparability, comprehensibility, definition of scale (interpretation), representation consistency (interpretation)*
4	context	*weighted average (context), impartial-contextual consistency, adaptivity*
5	aggregation	*aggregation consistency, aggregation, min or max operation, weighted average (aggregation), interpretation consistency (aggregation)*
6	cost	*cost/benefit, feasibility, acceptability, business relevance*

Group 1 comprises requirements stating that data quality metrics have to take values within a given range. *Simple ratio* and *representation consistency* aim at metric values in the range [0; 1]. *Measurability* results in a bounded range defined by the lowest and the highest discrete value. Hence, these requirements as well as *clarity of definition* (with respect to the range), *normalization* and *validity range* are assigned to this group. Group 2 contains requirements regarding the scale of measurement of the metric values. Since *definition of scale* may not only concern the interpretation of the metric values but also their scale, this requirement is included as well. Group 3 covers requirements claiming an interpretation of the metric values. Here, *clarity of definition* is interpreted as *interpretability*. In addition, metric values satisfying the *simple ratio* requirement can be interpreted as a percentage, and *interpretation consistency* requires a consistent semantic interpretation of the metric values regardless of the hierarchical level. While *comparability*, *comprehensibility* and *definition of scale* require some kind of interpretation of the metric values (e.g., as a percentage), *representation consistency* directly implies a clear interpretation with respect to the utility of the data under consideration. The requirements in Group 4 state that data quality metrics should be able to consider adequately the particular context of application, for example by means of weights that decrease or increase the influence of contextual characteristics. Group 5 concerns the (consistent) aggregation of the metric values on different data view levels. *Min or max operation* and *weighted average* specify how this aggregation has to be performed and *interpretation consistency* requires the same interpretation of the metric values on all data view levels. Finally, Group 6 focusses on the application of a data quality metric from a cost-benefit perspective. *Feasibility* is part of this group, because it requires that the costs for determining a metric's parameters are taken into account and that it should be possible to calculate the metric values in a widely automated way – a fact that results in lower application costs. *Business relevance* implies that a metric goes along with some benefit for the company, whereas *acceptability* is part of this group because business expectations are defined considering a cost-benefit perspective.

Table 1 provides an overview of the existing requirements for data quality metrics, which are partly fragmented and vaguely defined. Prior work does in fact lack a methodical framework and does not aim at stating and justifying which requirements for data quality metrics support decision-making under uncertainty and an economically oriented management of data quality. To address this research gap, in the next section we present a decision-oriented framework, enabling us to propose a set of requirements for data quality metrics in Section 4. In addition to that, the decision-oriented framework helps to clearly and unambiguously define the presented requirements as well as to justify them. In this way, it is possible to reason that a data quality metric should satisfy the presented requirements to support both decision-making under uncertainty and an economically oriented management of data quality. Finally, this set of clearly defined requirements combines, concretizes, and enhances the identified groups of existing requirements (cf. Table 1) and thus helps to alleviate the fragmentation within the literature on requirements for data quality metrics.

3 DECISION-ORIENTED FRAMEWORK

The decision-oriented framework for our work is based on the following fields: i) decision-making under uncertainty by considering the influence of assessed data quality metric values and ii) economically oriented management of data quality by considering the costs and benefits of applying data quality metrics.[1]

The literature on decision-making under uncertainty (and in particular under risk) uses the well-known concept of decision matrices to represent the situation decision makers are facing [Nitzsch 2006; Laux 2007; Peterson 2009]. Decision makers can choose among a number of alternatives while the corresponding payoff depends on the state of nature. Each possible state of nature occurs with a certain probability. Hence, in case of a risk-neutral decision maker (if this is not the case, the payoffs need to be determined considering risk adjustments), the one alternative is chosen which results in the highest expected payoff when considering the probability distribution over all possible states of nature. Table 2 illustrates a decision matrix for a simple situation with two alternatives a_i $(i = 1, 2)$, two possible states of nature s_j $(j = 1, 2)$, and the respective payoff p_{ij} for each pair (a_i, s_j). The probabilities of occurrence of the possible states of nature are represented by $w(s_j)$. To select the alternative with the highest expected payoff, the decision maker has to compare the expected payoffs for choosing alternative a_1 (i.e., $p_{11}w(s_1) + p_{12}w(s_2)$) and alternative a_2 (i.e., $p_{21}w(s_1) + p_{22}w(s_2)$). The two-by-two matrix serves for illustration purposes only. Generally, we represent the possible states of nature s_j $(j = 1, \dots, n)$ by the vector $S = (s_1, s_2, \dots, s_n)$, the respective probabilities of occurrence by $w(s_j)$, the alternatives a_i $(i = 1, \dots, m)$ by the vector $A = (a_1, a_2, \dots, a_m)$[2], and the payoffs for alternative a_i by the vector $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$. The *expected* payoff for choosing alternative a_i is denoted by $E(a_i, P_i, S) = \sum_{j=1}^{n} p_{ij}w(s_j)$; the maximum expected payoff is given by $max_{a_i}E(a_i, P_i, S)$. An overview of the notation is provided in Appendix A.

Table 2. Decision Matrix

	Probability $w(s_1)$	Probability $w(s_2)$
	State s_1	State s_2
Alternative a_1	Payoff p_{11}	Payoff p_{12}
Alternative a_2	Payoff p_{21}	Payoff p_{22}

Requirements for data quality metrics must guarantee that i) the metric values can support decision-making under uncertainty. To address i) it is necessary to examine the influence of data quality and thus of the data quality metric values on the components of the decision matrix (i.e., the *probabilities of occurrence,* the *payoffs,* and the *alternatives*). In this respect, literature provides useful insights. Heinrich et al. [2012], for example, propose a metric for the data quality dimension currency [cf. also Heinrich and Klier 2015]. The metric values represent probabilities that the data values under consideration still correspond to their real-world value at the instant of assessing data quality. They apply the metric to determine the *probabilities of occurrence* (represented by the metric values) in a decision situation. The influence of data quality on the *payoffs* is considered, for example, by Ballou et al. [1998], Even and Shankaranarayanan [2007], and Cappiello and Comuzzi [2009]. All of them argue that less than perfect data quality (represented by the data quality metric values) may affect and reduce the payoffs. Other works such as Fisher et al. [2003], Heinrich et al. [2007], and Jiang et al. [2007] examine the influence of data quality on the choice of the *alternative.*

More precisely, there are several possible ways to express, quantify and integrate the influence of data quality on decision-making. For instance, Even and Shankaranarayanan [2007] consider the effects of data quality on the payoffs for each record of a dataset. They select a subset of attributes which is relevant in the considered application scenario and set the payoffs for a record to zero if the value of at least one relevant attribute is missing. Moreover, having determined the influence of each data quality dimension, there may

[1] Note that i) may also be seen as an important means for ii). However, due to the high relevance of i) in the context of data quality metrics, we have decided to distinguish both cases.

[2] In case of a continuous decision space, this will be a vector of infinitely many alternatives. If not all alternatives are known, the concept of bounded rationality is applied [Simon 1956, 1969; Jones 1999].

be several ways to weight and aggregate these influences (e.g., by calculating the weighted sum across all data quality dimensions; cf. [Cappiello and Comuzzi 2009]). Therefore, we do not present an explicit formula or method to quantify the influence of data quality on the decision matrix but, instead, specify this impact more generally as follows: Let DQ represent the data quality metric value and $E(a_i, DQ, P_i, S)$ the *expected* payoff for choosing alternative a_i when considering DQ as well as the payoff vector P_i and the vector of states of nature S. Let further $max_{a_i} E(a_i, DQ, P_i, S)$ be the maximum *expected* payoff when considering data quality. It is obvious and in line with prior works (cf. above) that considering data quality may result in choosing a different optimal alternative as compared to not considering data quality (i.e., $a_1 = \underset{a_i}{argmax} E(a_i, DQ, P_i, S)$ and $a_2 = \underset{a_i}{argmax} E(a_i, P_i, S)$ with $a_1 \neq a_2$). Hence, it is useful to consider data quality by means of well-founded metrics in decision-making under uncertainty.

When developing requirements for data quality metrics, it is further necessary to take into account the field of ii) economically oriented management of data quality to avoid inefficient or impractical metrics. Existing literature has already addressed the question of whether to apply data quality improvement measures from a cost-benefit perspective [Campanella 1999; Feigenbaum 2004; Heinrich et al. 2007; Heinrich et al. 2012]. Indeed, applying data quality improvement measures may increase the data quality level and thus bring benefits. At the same time, the associated costs have to be taken into account and the improvement measures should only be applied if the benefits outweigh these costs. In decision-making, the benefits result from being enabled to choose a better alternative (i.e., with an additional expected payoff) due to the improved data quality. The costs include the ones for conducting the improvement measures as well as the ones for assessing data quality by means of data quality metrics. The latter have rarely been considered in the literature, even so they play an important role [Heinrich et al. 2007] and must not be neglected. Indeed, if applying a data quality metric is too resource-intensive, it may not be reasonable to do so from a cost-benefit perspective. Thus, requirements for data quality metrics have to explicitly consider this aspect.

Based on the literature on i) and ii) and the above discussion, Figure 1 presents the decision-oriented framework which is used to justify our requirements [for a similar illustration cf. Heinrich et al. 2007, 2009]. Data quality metrics are applied to data views to assess the data quality level (cf. I-III). The assessed data quality level (represented by the metric values) influences i) decision-making under uncertainty and in particular the chosen alternative, and the expected payoff of the decision maker (cf. IV-VI). Thus, the decision maker may apply improvement measures to increase the data quality level represented by the metric values (cf. IX). However, applying data quality improvement measures creates costs (cf. VII). This also holds for the application of the metric including the determination of its parameters (cf. II). Hence, the optimal data quality level (cf. VIII) has to be determined based on an economical perspective.

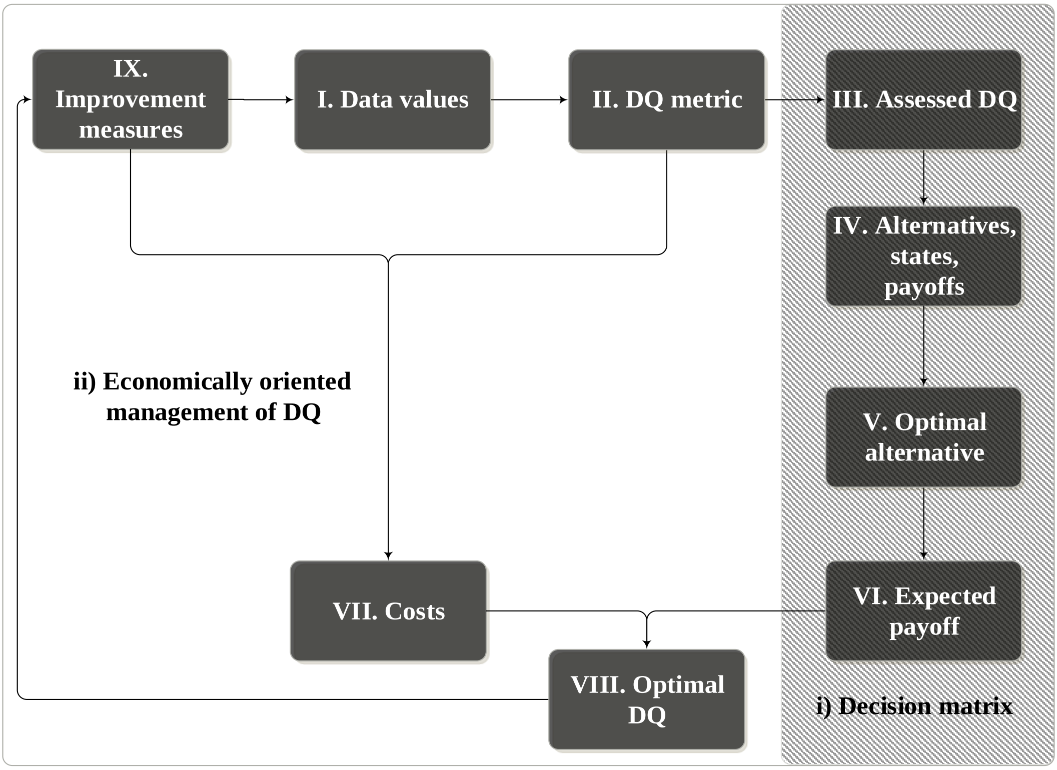

Figure 1. Decision-oriented Framework

4 REQUIREMENTS FOR DATA QUALITY METRICS

In this section, we present a set of five clearly defined requirements for data quality metrics. They combine, concretize, and enhance existing approaches covering the six groups of requirements identified in Section 2. Moreover, based on the decision-oriented framework we justify that our requirements support both i) decision-making under uncertainty and ii) an economically oriented management of data quality.

4.1 Requirement 1 (R1): Existence of Minimum and Maximum Metric Values

Group 1 states that data quality metrics have to take values within a given range. Most of the requirements in this group (e.g., *validity range* and *clarity of definition*) are vaguely defined and thus difficult to verify. Hence, both the relevance of these requirements and the possible consequences of them not being fulfilled remain unclear (e.g., *measurability* just claims that the range should be discrete). To address these issues, we propose and justify the following requirement:

Requirement 1 (R1) (Existence of minimum and maximum metric values). The metric values have to be bounded from below and from above and must be able to attain both a minimum (representing perfectly poor data quality) and a maximum (representing perfectly good data quality). In particular, for each real-world value ω_m, minimum and maximum value have to be attainable in regard to ω_m.

Justification. In a first step, we discuss the following statement (a) which will be used recurrently in the remainder of this justification:

(a) *There has to be exactly one metric value representing perfectly good data quality and exactly one metric value representing perfectly poor data quality.*

Re (a): Based on the definition of data quality by Orr [1998] used in this paper, perfectly good data quality implies a perfect agreement between stored data views and the real-world. This is a unique situation and therefore there is exactly one level of perfectly *good* data quality. In the case of the data quality dimension accuracy, existing metrics use a distance function to measure the difference between the real-world data values and the stored data values. Due to the finite number of possibilities for the stored data values (e.g., a 32 bit integer in Java can represent one of 2^{32}=4,294,967,296 possible numbers; this holds for other data types used for the assessed data value as well), there is always one or more data value(s) for which the distance to the real-world data value is maximal. For this/these data value/s, the data quality level "perfectly poor data" is reached and cannot become even worse; "even more inaccurate data" cannot be represented. Hence, there is exactly one level of perfectly poor data quality. Summing up and with respect to the discussion of Figure 1, as each data quality level is represented by a metric value and different metric values represent different data quality levels, there has to be exactly one metric value representing perfectly good data quality as well as exactly one metric value representing perfectly poor data quality.

Based on statement (a), we justify (R1). If a data quality metric does not fulfill (R1), this implies that the metric values

(b) are not bounded from below and/or from above and/or

(c) do not attain their minimum and/or maximum.

We denote by ω a stored data value (e.g., a stored customer address) of perfectly good data quality that perfectly represents the corresponding real-world value ω_m. Further, we denote the metric value for ω by $DQ(\omega, \omega_m)$.

Re (b): If there is no upper bound for the metric values, another stored data value ω' can exist which – compared to ω – results in a higher metric value (i.e., $DQ(\omega', \omega_m) > DQ(\omega, \omega_m)$ for the real-world value ω_m corresponding to ω and ω'). As higher metric values represent better data quality, this implies that ω' is of better data quality than ω. However, ω was defined to be of perfectly good data quality and only one metric value can represent perfectly good data quality (cf. statement (a)). Hence, the metric values indeed need to be bounded from above. The existence of a lower bound can be justified analogously by using a data value of perfectly poor data quality (e.g., the value 'NULL' stored for an unknown customer address which, however, does exist in the real-world).

Re (c): The metric values need to be bounded from below and from above (cf. re (b)). Hence, a supremum M (lowest upper bound) exists. If the metric values do not attain a maximum, it follows that $DQ(\omega, \omega_m) < M$ for a data value ω of perfectly good data quality. As M is the lowest upper bound, there exists another data value ω'' corresponding to the real-world value ω_m with $DQ(\omega, \omega_m) < DQ(\omega'', \omega_m) < M$ (otherwise, $DQ(\omega, \omega_m)$ would be an upper bound and the maximum of the metric values). However, ω was defined to be of perfectly good data quality. Hence, the metric values indeed have to attain a maximum. The existence of a minimum can be justified analogously by using a data value of perfectly poor data quality.

So far, we discussed the existence of a maximum (representing perfectly good data quality) and a minimum (representing perfectly poor data quality) for the metric values with regard to an arbitrary, but fixed real-world value ω_m. However, as there is always exactly one metric value representing perfectly good (resp. poor) data quality (cf. (a)), these maxima and minima coincide across all real-world values. Therefore, the metric values have to be bounded from below and from above and must attain both a minimum and a maximum (cf. I-III in Figure 1), equal for all real-world values.

When a data quality metric is represented by a mathematical function, (R1) means that this function has to be bounded from below and from above and must attain a minimum and maximum. However, some existing metrics [cf., e.g., Hipp et al. 2001; 2007; Hinrichs 2002; Alpar and Winkelsträter 2014] do not attain a minimum or maximum and may thus lead to a wrong evaluation of decision alternatives (cf. III-VI in Figure 1). In these cases it is, for example, not possible to decide whether the assessed data quality level can or should be increased to allow for better decision-making (cf. VI-IX in Figure 1). As a result, for instance, unnecessary improvement measures for data values of already perfectly good data quality may be performed since the metric values cannot represent the fact that perfectly good data quality has already been reached. Moreover, when assessing data quality multiple times with a metric which does not satisfy (R1), neither the comparability nor the validation (e.g., against a benchmark, such as a required completeness level of 90% of the considered database) of the metric values in different assessments are guaranteed. Moreover, when a specific data quality improvement measure is performed, no benchmark in the sense of a minimum and maximum exists to compare the rankings in the course of time (e.g., consider a user survey regarding the existing data quality level without any information in regard to the scale of values to be entered by the users). This contradicts an economically oriented management of data quality.

4.2 Requirement 2 (R2): Interval-Scaled Metric Values

The requirements in Group 2 focus on the scale of measurement of the metric values. These requirements have not been justified, and some of them do not specify a precise scale (e.g., *definition of scale* is not defined, but only illustrated by a very wide range of examples). To address this gap, we state and justify the following requirement:

Requirement 2 (R2) (Interval-scaled metric values). The values of a data quality metric have to be interval-scaled[3]. Based on the classification of scales of measurement [Stevens 1946], this means that differences and intervals can be determined and are meaningful.

Justification. We argue that a metric which does not provide interval-scaled values (cf. I-III in Figure 1) cannot support both the evaluation of decision alternatives and an economically oriented management of data quality in a well-founded way (cf. Section 3). For this, we take into account the decision matrix in Table 2 with the payoff vectors $P_1 = (p_{11}, p_{12})$ and $P_2 = (p_{21}, p_{22})$ for the alternatives a_1 and a_2 and let the expected payoffs for these alternatives be calculated based on the metric values DQ_1 and DQ_2, respectively. We consider a situation in which the expected payoffs for choosing alternative a_1 and alternative a_2 are the same (i.e., $E(a_1, DQ_1, P_1, S) = E(a_2, DQ_2, P_2, S)$) while $p_{11} > p_{21}$, $p_{12} = p_{22}$, and $DQ_1 < DQ_2$ holds. Hence, the decision maker faces a situation in which in state s_1 choosing alternative a_1 goes along with a higher payoff than choosing a_2 ($p_{11} > p_{21}$), but due to the lower metric value DQ_1 compared to DQ_2, the expected payoff for both alternatives which takes into account the effects of DQ_1 and DQ_2 is the same (cf. III-VI in Figure 1). In this situation, the decision maker is indifferent between the two alternatives[4]. Thus, the lower payoff for a_2 – compared to a_1 – is accepted if its estimation is based on data of higher data quality. This means that the decision maker equally evaluates both a change in payoffs from p_{11} to p_{21} and a change in data quality metric values from DQ_1 to DQ_2. As both the payoffs and expected payoffs are interval-scaled, the differences between payoffs (resp. expected payoffs) are meaningful and their change can be quantified and evaluated by calculating these differences. To support decision-making under uncertainty, this quantified, interval-scaled change in payoffs has to be comparable to a change in data quality. Hence, it has to be possible to calculate the change between the metric values DQ_1 and DQ_2. When the values provided by a metric are not interval-scaled, there is a missing interpretability of the changes between the metric values compared to the respective existing and meaningful differences in the payoffs which impedes the evaluation of decision alternatives. Hence, at most ordinal-scaled data quality metric values cannot support both the evaluation of decision alternatives and an economically oriented management of data quality.

(R2) has a significant practical impact. Indeed, many existing data quality metrics [cf., e.g., Ballou et al. 1998; Hinrichs 2002], which do not provide interval-scaled values, may lead to wrong decisions when evaluating different decision alternatives (cf. III-VI in Figure 1). Moreover, when evaluating, interpreting and comparing the effects of different data quality improvement measures for an economically oriented management of data quality, interval-scaled metric values are highly relevant. For example, let an ordinal-scaled metric take the values “very good”, “good”, “medium”, “poor” and “very poor”. Then there is no possibility of specifying the meaning of the difference between “very good” and “medium” and a decision maker cannot assess whether it would have the same business value as a difference in payoffs of \$500 or \$600. In contrast, this difference in payoffs may be equivalent to a difference of 0.2 in metric values for an interval-scaled metric. In particular, it is not enough to state which measure results in the greatest improvement of the data

[3] They may also be ratio-scaled, which is a stronger property and includes interval scaling [Stevens 1946].

[4] If such a situation does not exist, the decision is trivial: If $E(a_1, DQ_1, P_1, S) > E(a_2, DQ_2, P_2, S)$ holds for $p_{11} > p_{21}$, $p_{12} = p_{22}$ and all possible values for DQ_1 and DQ_2 (i.e., it is not necessarily $DQ_1 < DQ_2$), the decision maker will always choose a_1 regardless of the metric values. In this case, data quality does not matter, which means that assessing data quality is not necessary at all. The same argumentation applies analogously for $E(a_1, DQ_1, P_1, S) < E(a_2, DQ_2, P_2, S)$ where alternative a_2 will always be chosen.

quality level based on ordinal-scaled metric values. In the example of an ordinal-scaled metric above, it cannot be determined whether an improvement from "very poor" to "medium" is of the same magnitude as an improvement from "medium" to "very good". Similarly, it is unclear whether an improvement from "very poor" to "medium" is twice as much as an improvement from "very poor" to "poor". In contrast, for an interval-scaled metric, an improvement of 0.2 is twice as much as an improvement of 0.1. To ensure the selection of efficient data quality improvement measures, their benefits (i.e., the additional expected payoff) resulting from a clearly specified increase in the data quality level need to be determined precisely and compared to their costs (cf. VI-IX in Figure 1).

The requirements in Group 3 state that the metric values must have an interpretation. However, existing requirements (e.g., *comprehensibility*, *comparability*, *interpretability*, *definition of scale*, *interpretation consistency*, and *clarity of definition*) have neither been justified nor do they specify what exactly is meant by interpretation, making the verification of data quality metrics in this regard very difficult. In the following, we argue that we do not need to define a separate requirement for Group 3, because a clear interpretation is already ensured by the combination of (R1) and (R2). Indeed, a metric which meets both (R1) and (R2) is *interpretable* in terms of the measurement unit *one* [Bureau International des Poids et Mesures 2006]. To justify this, let m be the minimum (representing perfectly poor data quality) and M be the maximum (representing perfectly good data quality) of the metric values (cf. (R1)). Since equal differences result in equidistant numbers on an interval scale (cf. (R2)), each value DQ of the metric can be interpreted as the $\frac{(DQ-m)}{(M-m)}$ fraction of the maximum difference $(M - m)$. Thus, a data quality metric that meets both (R1) and (R2) is inherently interpretable in terms of the measurement unit *one* (i.e., as percentage).

A clear interpretation of the metric values is helpful to understand the actual meaning of the data quality level and is thus important in practical applications, such as the communication to business users. This is the case if the metric values are ratio-scaled. Ratio-scaled metric values support statements such as "a metric value of 0.6 is twice as high as a metric value of 0.3". Ratio-scale can be achieved by a simple transformation of each interval-scaled data quality metric whose minimum m of the metric values is transformed to 0 so that each metric value can be interpreted as a fraction with respect to the maximum data quality value.

4.3 Requirement 3 (R3): Quality of the Configuration Parameters and the Determination of the Metric Values

Group 4 contains requirements stating that it must be possible to adjust a data quality metric to adequately reflect the particular context of application. This, however, addresses only one relevant aspect. There are well-known scientific quality criteria (i.e., objectivity, reliability, and validity) that must be satisfied by data quality metrics but have not been considered in the literature yet. In addition, not only the metric values, but also the configuration parameters of a data quality metric should satisfy these quality criteria to avoid inadequate results (cf. II-III in Figure 1).[5] To address these drawbacks, we propose and justify the following requirement:

Requirement 3 (R3) (Quality of the configuration parameters and the determination of the metric values). It must be possible to determine the configuration parameters of a data quality metric according to the quality criteria objectivity, reliability, and validity [cf. Allen and Yen 2002; Cozby and Bates 2012; Zikmund et al. 2012]. The same holds for the determination of the metric values.

There exists a large body of literature dealing with the quality criteria *objectivity*, *reliability,* and *validity* of measurements in general [cf., e.g., Litwin 1995; Allen and Yen 2002; Marsden and Wright 2010; Cozby and Bates 2012; Zikmund et al. 2012]. In the following, we first briefly discuss these criteria for the context of data quality metrics. Afterwards, we justify their relevance based on our decision-oriented framework.

Objectivity of both the configuration parameters and the data quality metric values denotes the degree to which the respective parameters and values as well as the procedures for determining them (e.g., SQL queries) are independent of external influences (e.g., interviewers). This criterion is especially important for

[5] Note that in line with our focus on a methodical perspective on requirements for data quality metrics, we concentrate on methodical criteria. Organizational aspects such as the frequency of applying the metric (defined and idiosyncratic per company) are not discussed.

data quality metrics requiring expert estimations to determine the configuration parameters or the metric values [cf., e.g., Ballou et al. 1998; Hinrichs 2002; Cai and Ziad 2003; Even and Shankaranarayanan 2007; Hüner et al. 2011; Heinrich and Hristova 2014]. Here, *objectivity* is violated if the estimations are provided by too few experts or if external influences such as the particular behavior of the interviewers are not minimized. In general, *objectivity* becomes an issue if metrics lack a precise specification of (sound) procedures for the determination of the respective parameters and values. In this case, metrics may result in different results if applied multiple times. To avoid highly subjective results and ensure *objectivity*, the data quality metric and its configuration parameters have to be unambiguously (e.g., formally) defined and determined with objective procedures (e.g., statistical methods; cf., e.g., [Heinrich et al. 2012]).

Reliability of measurement refers to the accuracy with which a parameter is determined. *Reliability* conceptualizes the replicability of the results of the methods used for the determination of the configuration parameters or the metric values. In particular, methods will not be reliable, if expert estimations which change over time or among different groups of experts are involved. *Reliability* can be analyzed based on the correlation of the results obtained from the different measurements. Thus, data quality metrics which rely on expert estimations [cf., e.g., Ballou et al. 1998; Even and Shankaranarayanan 2007; Hüner et al. 2011; Heinrich and Hristova 2014] have to define a reliable procedure to determine the configuration parameters and the metric values. Generally, to ensure *reliability* of the configuration parameters and the metric values, correct database queries or statistical methods may be used. In this case, the result of the respective procedure remains the same when being applied multiple times to the same data.

Validity is defined as the degree to which a metric "measures what it purports to measure" [Allen and Yen 2002] or as "the extent to which [a metric] is measuring the theoretical construct of interest" [Marsden and Wright 2010]. Hence, the *validity* of a method for determining the configuration parameters or the metric values refers to the degree of accuracy with which a proposed method actually measures what it should measure.[6] Typically, the *validity* of the determination of a configuration parameter or a metric value is violated if the determination contradicts the aim. There are several examples which illustrate the practical relevance of *validity* in the context of data quality metrics. The metric for timeliness by Batini and Scannapieco [2006, p. 29], for example, involves the configuration parameter $Currency$ which is intended to represent "how promptly data are updated". Its mathematical specification $Currency = Age + (DeliveryTime - InputTime)$, however, seems to contradict this aim. Similarly, Hüner et al. [2011, p. 150] state that a metric value of zero indicates that "each data object validated contains at least one critical defect". However, the mathematical definition of the metric reveals that, to be zero, each data object must actually contain *all* possible critical defects. *Validity* can be achieved by consistent definitions, database queries, or statistical estimations constructed to determine the corresponding parameter or value according to its definition. Additionally, restricting the application domain of a metric [cf., e.g., Ballou et al. 1998; Heinrich et al. 2007] also contributes to *validity*.

Justification. To justify the relevance of (R3) based on the decision-oriented framework, we consider a data quality metric for which *objectivity, reliability* and/or *validity* are violated but their values are used to support decision-making under uncertainty (cf. Figure 1). For example, *objectivity* and/or *reliability* may be violated due to different expert estimations for the configuration parameters of the metric and *validity* may be violated due to an inaccurate definition of the metric or its configuration parameters (cf. above). We analyze a decision situation as illustrated by the decision matrix in Table 2. In case *objectivity* and/or *reliability* are violated, two applications of the data quality metric result in two different data quality levels DQ_1 and DQ_2 with $DQ_1 \neq DQ_2$ (e.g., depending on different expert estimations). In case *validity* is violated, the data quality level DQ_1 estimated by the metric does not accurately represent the actual data quality level DQ_2 in the real-world. In either case, consider that DQ_1 and DQ_2 result in choosing different alternatives. To be more precise, this means $a_1 = argmax_{a_i} E(a_i, DQ_1, P_i, S)$ and $a_2 = argmax_{a_i} E(a_i, DQ_2, P_i, S)$ with $a_1 \neq a_2$[7] (cf. III-VI in Figure 1). If *objectivity* and/or *reliability* are violated, it is not clear to the decision maker whether DQ_1, DQ_2 or none of them correctly reflects the actual data quality level and thus whether a_1 or a_2 is the accurate decision. Similarly, if *validity* is violated, then the decision maker will choose a_1

[6] If validity and reliability are fulfilled for a data quality metric, variations in metric values reflect variations in the level of data quality (i.e., sensitivity is guaranteed; cf., e.g., [Allen and Yen 2002]).

[7] In case $a_1 = a_2$, data quality does not matter, which means that assessing data quality is not necessary at all (cf. the justification of (R2)).

instead of the accurate choice a_2. Thus, in case *objectivity*, *reliability* and/or *validity* are violated, decision makers will make wrong decisions.

The above justification reveals that data quality metrics which do not fulfill (R3) can lead to wrong decisions when evaluating alternatives (cf. III-VI in Figure 1). In addition, such metrics result in serious problems when evaluating data quality improvement measures (cf. VII-IX in Figure 1). Indeed, assessing data quality before and after a data quality improvement measure with a metric not fulfilling (R3) results in inaccurate metric values. This makes it impossible to determine the increase in the data quality level in a well-founded way (e.g., a data quality improvement measure evaluated as effective before its application may not even lead to an increase in the actual data quality level). To support an economically oriented management of data quality, it is thus important to ensure (R3).

4.4 Requirement 4 (R4): Sound Aggregation of the Metric Values

Group 5 addresses the consistent aggregation of the metric values on different data view levels. Again, the requirements in this group are not motivated based on a sound framework. In addition, applying the *min or max* and the *weighted average operations* – as proposed by existing works – does not necessarily assure a consistent aggregation. We address these issues by the following requirement:

Requirement 4 (R4) (Sound aggregation of the metric values). A data quality metric has to be applicable to single data values as well as to sets of data values (e.g., tuples, relations, and a whole database). Furthermore, it has to be assured that the aggregation of the resulting metric values is consistent throughout all levels. To be more precise, for data D_{l+1} at a data view level $l+1$ with a disjoint decomposition $D_{l+1} = D_l^1 \cup D_l^2 \cup \ldots \cup D_l^H$ at data view level l (i.e., $D_l^i \cap D_l^j = \emptyset$ for all $i, j \in \{1, \ldots, H\}, i \neq j$), there has to exist an aggregation function f: $DQ(D_{l+1}) = f(DQ(D_l^1), \ldots, DQ(D_l^H))$ with $f(DQ(D_l^1), \ldots, DQ(D_l^H)) = f(DQ(\widetilde{D}_l^1), \ldots, DQ(\widetilde{D}_l^K))$ for all disjoint decompositions D_l^h, $\widetilde{D}_l^k$ of D_{l+1}.

Justification. To justify the relevance of (R4), we argue that a data quality metric needs to

(a) *be applicable to different data view levels and*
(b) *provide a consistent aggregation of metric values*

in order to support decision-making under uncertainty and an economically efficient management of data quality.

Re (a): Consider a situation (cf. Figure 1) in which data used for decision-making is not restricted to the level of single data values, but also covers sets of data values (e.g., tuples, relations, and the whole database). This implies that for decision-making under uncertainty and an economically oriented management of data quality, it must be possible to determine data quality at several data view levels.

Re (b): Consider a data quality metric which is defined at both a lower data view level l (e.g., relations) and a higher data view level $l+1$ (e.g., database). In the following, we justify that the metric values must have a consistent aggregation from l to $l+1$. To be more precise, we argue that if an aggregation function f for determining the metric value at level $l+1$ based on the metric values at level l does not assure a consistent aggregation, the metric values cannot support decision-making under uncertainty and an economically oriented management of data quality in a well-founded way (cf. Section 3). In this case, the aggregation of the metric values at l to the metric value at $l+1$ does not adequately reflect the characteristics of the underlying datasets at l (e.g., size, importance). For our argumentation, we consider a disjoint decomposition of a dataset D_{l+1} at $l+1$ into the subsets D_l^h $(h = 1, \ldots, H)$ at l (e.g., a database D_{l+1} which is decomposed into non-overlapping relations D_l^h): $D_{l+1} = D_l^1 \cup D_l^2 \cup \ldots \cup D_l^H$ and $D_l^i \cap D_l^j = \emptyset \ \forall i \neq j$. The metric values for the subsets D_l^h are denoted by $DQ(D_l^h)$. On this basis, the metric value for D_{l+1} can be determined by means of the aggregation function f: $DQ(D_{l+1}) = f(DQ(D_l^1), \ldots, DQ(D_l^H))$. If the aggregation function f does not assure a consistent aggregation of the metric values from l to $l+1$, there exists another decomposition $D_{l+1} = \widetilde{D}_l^1 \cup \widetilde{D}_l^2 \cup \ldots \cup \widetilde{D}_l^K$ of D_{l+1} at l with $DQ'(D_{l+1}) = f(DQ(\widetilde{D}_l^1), \ldots, DQ(\widetilde{D}_l^K))$, $\widetilde{D}_l^i \cap \widetilde{D}_l^j = \emptyset \ \forall i \neq j$ and $DQ'(D_{l+1}) \neq DQ(D_{l+1})$. Following this, the resulting metric value for D_{l+1} depends on the decomposition of the dataset and can hence be manipulated accordingly (i.e., there are two or more possible metric values for the same dataset). Thus, we face the same situation as in the justification of

(R3) where it is also not known which metric value actually represents the "real" data quality level of the dataset D_{l+1}. It analogously follows that this situation results in wrong decisions (cf. III-VI in Figure 1). To sum up, a data quality metric requires a consistent aggregation of the metric values throughout the different data view levels to support decision-making under uncertainty and an economically oriented management of data quality (cf. Section 3).

When data quality metrics are seen as mathematical functions, (R4) means that these functions for the different data view levels have to be compatible with aggregation. Decision situations usually rely on the data quality of (large) sets of data values. However, many data quality metrics in the literature do not provide (consistent) aggregation rules for different data view levels [cf., e.g., Hipp et al. 2001; Hipp et al. 2007; Li et al. 2012; Alpar and Winkelsträter 2014]. As the above justification reveals, this may lead to wrong decisions when evaluating different decision alternatives (cf. III-VI in Figure 1). In addition, a consistent interpretation of the metric values on all aggregation levels is important to support an economically oriented management of data quality. Otherwise, (repeated) measurements of data quality will provide inconsistent and/or wrong results (e.g., when assessing sets of data values that change their volume over time), making it impossible to precisely determine the benefits of improvement measures and to decide whether they should be applied from a cost-benefit perspective (cf. VI-IX in Figure 1).

4.5 Requirement 5 (R5): Economic Efficiency of the Metric

Finally, Group 6 comprises requirements addressing the cost-benefit perspective when applying data quality metrics[8]. Existing requirements in this group are not motivated based on a framework. Moreover, for some of them, their definition, specification, and interpretation remain unclear (e.g., *business relevance* and how to determine the threshold for *acceptability*), making them difficult to verify. We address these issues by proposing and justifying the following requirement):

Requirement 5 (R5) (Economic efficiency of the metric). The configuration and application of a data quality metric have to be efficient from an economic perspective. In particular, the additional expected payoff from the intended application of a metric has to outweigh the expected costs for determining both the configuration parameters and the metric values.

Justification. To justify (R5), we analyze a decision situation as shown in the decision matrix in Table 2. Let alternative a_1 be chosen by a decision maker who does not consider data quality in decision-making (and thus does not apply the metric). Furthermore, let another alternative a_2 be chosen if data quality is considered. To be more precise, it holds $a_1 = argmax_{a_i} E(a_i, P_i, S)$ and $a_2 = argmax_{a_i} E(a_i, DQ, P_i, S)$ with $a_1 \neq a_2$.[9] In this situation, from a decision-making perspective, considering data quality represents *additional information* influencing the evaluation of the decision alternatives and their choice. This means that the existing data quality level is an additional information affecting the (ex post) realized payoffs. Thus, the benefit of this additional information is assessed by the difference between the expected payoffs (cf. III-VI in Figure 1) when choosing a_1 resp. a_2 both under consideration of the additional information, which means, $E(a_2, DQ, P_2, S) - E(a_1, DQ, P_1, S)$ [for details cf. Heinrich and Hristova 2016]. Thereby, the application of the data quality metric is economically efficient and therefore justifiable with respect to the decision-oriented framework (cf. Figure 1) if and only if the difference between the expected payoffs outweighs the expected costs for applying the data quality metric. Otherwise, the metric contradicts an economically oriented management of data quality.

Regarding (R5), especially metrics requiring configuration parameters not directly available to the user have to be analyzed in detail. For example, the metric for correctness by Hinrichs [2002] involves determining real-world values as input, which is usually very resource-intensive and raises the question why a subsequent data quality assessment is even necessary (for a detailed discussion cf. Section 5.4). In case of metrics not fulfilling (R5), the determination of the configuration parameters or the procedure for determining the metric values is expected to be too costly compared to the estimated additional expected payoffs (cf. I-IX in Figure 1). In some cases, it may be possible to use automated approximations and estimations (especially for

[8] We consider a decision scenario (and the related expected payoffs and costs) in which a data quality metric supports an economically oriented management of data quality from a methodical perspective. We do not focus on organizational aspects such as the conduction of a decision-making process in organizations.

[9] In case $a_1 = a_2$, data quality does not matter, which means that assessing data quality is not necessary at all (cf. the justification of (R2)).

configuration parameters) to reduce the effort. Metrics not fulfilling (R5) can still be valuable from a theoretical perspective, but they are not of practical relevance. (R5) is of particular importance in data governance and data quality management. Indeed, metrics not fulfilling (R5) are usually not suitable for use in a data governance initiative for data quality assessment, as the valuation and success of actions (such as applying a data quality metric) taken in such initiatives is ultimately to be determined by economic efficiency [Sarsfield 2009].

5 APPLICATION OF THE REQUIREMENTS

We demonstrate the applicability and efficacy of our requirements by evaluating five metrics from literature [Ballou et al. 1998; Hinrichs 2002; Blake and Mangiameli 2011; Yang et al. 2013; Alpar and Winkelsträter 2014]. We chose these metrics covering timeliness, completeness, reliability, correctness and consistency to provide a broad perspective on different dimensions of data quality and to show that the presented requirements can indeed be applied to various dimensions for data views and data values stored in an information system. To make the evaluation of the metrics more transparent and comprehensible, we refer to the following context of application [cf. Even et al. 2010; Heinrich and Klier 2015]: Based on the stored data of existing customers (e.g., corporate customers), a company has to decide which customers to contact with a new product offer in a CRM mailing campaign. The two decision alternatives for the company with respect to each customer in the database are a_1: to select the customer for the campaign or a_2: not to do so. The possible states of nature (occurring depending on a certain probability of acceptance) are s_1: the customer accepts or s_2: the customer rejects the offer. The benefits of applying a data quality metric in this context are generally non-negligible. Indeed, considering the quality of the customer data (as discussed by [Even et al. 2010] and [Heinrich and Klier 2015]) will lead to better decisions (e.g., if an offer is sent to an outdated or incomplete address, this will only cause mailing costs).

5.1 Metric for Timeliness by Ballou et al. [1998]

The data quality metric for timeliness proposed by Ballou et al. [1998] is defined as follows:

$$Timeliness := \max\left[1 - \frac{age\ of\ the\ data\ value}{shelf\ life}, 0\right]^{s} \tag{1}$$

The parameter *age of the data value* represents the time difference between the occurrence of the real-world event (i.e., when the data value was created in the real-world) and the assessment of timeliness of the data value. The parameter *shelf life* is defined as the maximum length of time the values of the considered attribute remain up-to-date. Thus, a higher value of the parameter *shelf life*, ceteris paribus, implies a higher value of the metric for timeliness, and vice versa. The exponent $s > 0$, which has to be determined based on expert estimations, influences the sensitivity of the metric to the ratio $\frac{age\ of\ the\ data\ value}{shelf\ life}$. In Table 3, we present the evaluation of the metric based on the requirements.

Table 3. Evaluation of the Metric by Ballou et al. [1998]

R1: Existence of minimum and maximum metric values	(Fulfilled)
For all values of the parameter $s > 0$, the metric values are within the bounded interval [0; 1]. The minimum of zero (which represents perfectly poor data quality) is attained if the parameter *age of the data value* is greater than or equal to the parameter *shelf life*. The maximum of one (which represents perfectly good data quality) is attained if the parameter *age of the data value* equals zero (e.g., a stored customer address is certainly up-to-date). It follows that (R1) is fulfilled.	

<table>
<tr><th>R2: Interval-scaled metric values</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">For $s = 1$ the metric values can be interpreted as the percentage of the data value's remaining shelf life (e.g., a stored customer address is up-to-date with 50%). As a consequence, for $s = 1$ we observe a ratio scale which implies that the values are interval-scaled as well. Apart from this particular case (i.e., for $s \neq 1$), however, the metric values are not interval-scaled. This is due to the fact that for any two interval scales it is always possible to transform one of them to the other by applying a positive linear transformation of the form $x \mapsto ax + b$ (with $a > 0$) [Allen and Yen 2002]. Obviously, such a transformation does not exist for $s \neq 1$, as the mapping $x \mapsto x^s$ is not linear for $s \neq 1$. That is why the metric values are generally not interval-scaled and (R2) is not fulfilled. To conclude: The parameter s allows to control the sensitivity of the metric values with respect to the ratio of <i>age of the data value</i> and <i>shelf life</i>, which may be advantageous in specific contexts. To obtain interval-scaled metric values, however, the value $s = 1$ has to be chosen.</td></tr>
<tr><th>R3: Quality of the configuration parameters and the determination of the metric values</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">In the context of corporate customer data, the values of the attribute "address" do not have a known and fixed maximum shelf life. Indeed, company addresses are not characterized by a maximum length of time during which they remain up-to-date (e.g., some companies have been located at the same address for hundreds of years). In this case, it is not possible to determine a fixed value for the configuration parameter <i>shelf life</i> of the metric. As a result, (R3) is not fulfilled.</td></tr>
<tr><th>R4: Sound aggregation of the metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The authors propose to use the weighted arithmetic mean to aggregate the metric values from single data values to a set of data values. (R4) is fulfilled, as this aggregation rule ensures a consistent aggregation of the metric values on all levels. This allows to use the results from an application of the metric for a broad variety of decisions. For example, in the context of customer data, the metric values can be used for the selection of individual customers for the mailing campaign (i.e., a decision on the level of tuples). However, the metric values could – after aggregation – also be used for the decision whether to perform a data quality improvement measure for a larger portfolio of customers.</td></tr>
<tr><th>R5: Economic efficiency of the metric</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">Ballou et al. [1998] define the parameter <i>age of the data value</i> based on the point of time when the data value was created in the real-world. Therefore, to determine the parameter <i>age of the data value</i> for the given context of a customer's address, it has to be known when the customer moved to this address. This point of time, however, is usually neither stored nor easily accessible for companies (e.g., due to privacy protection laws) making the expected costs of configuration parameter determination very high. Indeed, for the above context of a customer database it would not be efficient to determine the configuration parameter <i>age of the data value</i> to be able to calculate the metric values for the company's customers. Actually, it would even be easier and less resource-intensive – independent of the benefits of the campaign – to directly evaluate whether the data values are still up-to-date (e.g., by contacting the customers). Therefore, (R5) is not fulfilled in our considered application context.</td></tr>
</table>

Overall, while the metric for timeliness proposed by Ballou et al. [1998] fulfills (R1) and (R4), it does not fulfill (R2), (R3), and (R5).

5.2 Metric for Completeness by Blake and Mangiameli [2011]

The metric for completeness by Blake and Mangiameli [2011] is defined as follows. On the level of data values, a data value is incomplete (i.e., the metric value is zero) if and only if it is 'NULL', otherwise it is complete (i.e., the metric value is one). Here, all data values which represent missing or unknown values in a specific application scenario (e.g., blank spaces or '9/9/9999' as a date value) are represented by the data value 'NULL'. A tuple in a relation is defined as complete if and only if all data values are complete (i.e., none of its data values is 'NULL'). For a relation R, let T_R be the number of tuples in R which have at least one 'NULL'-value and let N_R be the total number of tuples in R. Then, the completeness of R is defined as follows:

$$Completeness \coloneqq 1 - \frac{T_R}{N_R} = \frac{N_R - T_R}{N_R} \tag{2}$$

The evaluation of the metric with respect to the requirements is presented in Table 4:

Table 4. Evaluation of the Metric by Blake and Mangiameli [2011]

<table>
<tr><th>R1: Existence of minimum and maximum metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The metric values are within the bounded interval [0; 1]. This holds for all aggregation levels. The minimum of zero (which represents perfectly poor data quality) on the level of data values, tuples, and relations is attained, if a data value equals 'NULL' (e.g., the street of a single customer address is not stored), if a tuple contains at least one data value which equals 'NULL', and if each tuple of a relation contains at least one data value which equals 'NULL', respectively. The maximum of one (which represents perfectly good data quality) on the level of data values, tuples, and relations is attained if a data value does not equal 'NULL', if a tuple does not contain any data value which equals 'NULL', and if a relation does not contain any tuple with data values which equal 'NULL', respectively. It directly follows that (R1) is fulfilled.</td></tr>
<tr><th>R2: Interval-scaled metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">On the levels of data values and tuples, the metric values are interval-scaled (i.e., the difference between the only two possible metric values zero and one is meaningful). On the level of relations, the metric values are defined as the percentage of tuples which do not contain any data value which equals 'NULL' (e.g., 50% of all tuples storing customer data are complete). That implies a ratio scale, and thus the values are also interval-scaled. Therefore, (R2) is fulfilled. Based on the metric values' interpretation, the impact of a data quality improvement measure can thus be assessed precisely. For instance, a change in metric values from 0.4 to 0.7 means that instead of 40%, now 70% of all tuples are complete, which may be important for an appropriate selection of customers.</td></tr>
<tr><th>R3: Quality of the configuration parameters and the determination of the metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">All configuration parameters of the metric (i.e., whether a data value equals 'NULL'; whether a tuple contains a data value, which equals 'NULL'; and the number of tuples in a relation and how many of them contain at least one data value, which equals 'NULL') can be determined by means of simple database queries. Hence, the quality criteria objectivity, reliability, and validity are fulfilled. The metric values can be determined by means of mathematical formulae in an objective and reliable way. As the metric quantifies the data quality dimension completeness at different levels according to the corresponding definition, the determination of the metric values is valid. To sum up, (R3) is fulfilled.</td></tr>
<tr><th>R4: Sound aggregation of the metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The metric is applicable to single data values as well as to sets of data values (tuples and relations). The determination of the metric values on the different aggregation levels follows well-defined rules allowing for a consistent aggregation. Therefore, (R4) is fulfilled.</td></tr>
<tr><th>R5: Economic efficiency of the metric</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The parameters of the metric can be determined by means of database queries and the metric values can be determined by means of mathematical formulae, both of them in an automated and effective way and at negligible costs. In case the benefits from applying the metric are non-negligible (cf. given context of application), the application of the metric is efficient and thus fulfills (R5). For instance, in the application context of the CRM mailing campaign, the costs for applying the metric will easily be made up for by saving costs for sending mailings in case of incomplete customer records.</td></tr>
</table>

Overall, the metric by Blake and Mangiameli [2011] satisfies all requirements (R1) to (R5).

5.3 Metric for Reliability by Yang et al. [2013]

The data quality metric for reliability proposed by Yang et al. [2013] is defined based on the answers to n equally important[10] questions referring to the reliability of a given dataset (e.g., a database). In particular,

[10] In the application of Yang et al. [2013], $n = 21$ is used. The authors also discuss the use of so-called "red criteria", which always need to be fulfilled. As their use is not decisive for the evaluation of the proposed metric, we do not further consider them here.

the answer to question i is represented by the triangular fuzzy number Q_i =(a_{1i}, a_{2i}, a_{3i}), where $a_{1i} = s_i c_i$, $a_{2i} = s_i$ and $a_{3i} = s_i c_i + 1 - c_i$ with $s_i \in [0;\ 1]$ being the satisfaction degree of question i and $c_i \in [0;\ 1]$ the corresponding certainty degree. The reliability of a dataset is defined by the total score:

$$Reliability \coloneqq \textstyle\sum_{i=1}^{n} Q_i \tag{3}$$

This reliability is then matched to one of three fuzzy sets, representing different levels of reliability. In order to evaluate this metric with regard to our requirements, we consider the approach proposed by the authors in a decision support context (such as the aforementioned CRM mailing campaign) to defuzzify the total score in (3). We apply the centroid method as the most common defuzzification approach [Driankov et al. 1996]. On this basis, given a triangular fuzzy number Q_i =(a_{1i}, a_{2i}, a_{3i}), the defuzzification operator is

$$C\colon (a_{1i}, a_{2i}, a_{3i}) \mapsto \frac{a_{1i} + a_{2i} + a_{3i}}{3} \tag{4}$$

and the defuzzified reliability of a dataset is defined by

$$\sum_{i=1}^{n} C(Q_i) \tag{5}$$

In Table 5, we present the evaluation of the metric based on the requirements.

Table 5. Evaluation of the Metric by Yang et al. [2013]

<table>
<tr><th>R1: Existence of minimum and maximum metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The maximum reliability is achieved if all n questions are assigned both a satisfaction degree and certainty degree of one (e.g., all experts are certain that customer information is fully reliable). In this case, the defuzzified score in (5) is n. The minimum reliability is achieved if all n questions are assigned a satisfaction degree of zero and a certainty degree of one (e.g., all experts are certain that customer information is not reliable at all). In this case, the defuzzified score in (5) is 0. Thus, (R1) is fulfilled.</td></tr>
<tr><th>R2: Interval-scaled metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">Consider two different reliability scores generated on two different datasets:
$$Score_1 = \left(q_1^{(1)}, q_2^{(1)}, q_3^{(1)}\right) \tag{6}$$
$$Score_2 = \left(q_1^{(2)}, q_2^{(2)}, q_3^{(2)}\right) \tag{7}$$
where $q_k^{(j)} = \sum_{i=1}^{n} a_{ki}^{(j)}, j \in \{1,2\}, k \in \{1,2,3\}$ and $(a_{1i}^{(j)}, a_{2i}^{(j)}, a_{3i}^{(j)})$ as defined above. Then, the defuzzified values of $Score_1$ and $Score_2$ are:
$$C(Score_1) = \frac{q_1^{(1)} + q_2^{(1)} + q_3^{(1)}}{3} \tag{8}$$
$$C(Score_2) = \frac{q_1^{(2)} + q_2^{(2)} + q_3^{(2)}}{3} \tag{9}$$
As a result:
$$\begin{aligned} C(Score_1) - C(Score_2) &= \frac{\left(q_1^{(1)} - q_1^{(2)}\right) + \left(q_2^{(1)} - q_2^{(2)}\right) + \left(q_3^{(1)} - q_3^{(2)}\right)}{3} \\ &= \textstyle\sum_{i=1}^{n} \frac{\left(a_{1i}^{(1)} - a_{1i}^{(2)}\right) + \left(a_{2i}^{(1)} - a_{2i}^{(2)}\right) + \left(a_{3i}^{(1)} - a_{3i}^{(2)}\right)}{3} \\ &= \textstyle\sum_{i=1}^{n} \frac{\left(s_i^{(1)} c_i^{(1)} - s_i^{(2)} c_i^{(2)}\right) + \left(s_i^{(1)} - s_i^{(2)}\right) + \left(s_i^{(1)} c_i^{(1)} + 1 - c_i^{(1)} - s_i^{(2)} c_i^{(2)} - 1 + c_i^{(2)}\right)}{3} \\ &= \textstyle\sum_{i=1}^{n} \frac{2\left(s_i^{(1)} c_i^{(1)} - s_i^{(2)} c_i^{(2)}\right) + \left(s_i^{(1)} - s_i^{(2)}\right) + \left(c_i^{(2)} - c_i^{(1)}\right)}{3} \\ &= \textstyle\sum_{i=1}^{n} \frac{\left(2 s_i^{(1)} c_i^{(1)} + s_i^{(1)} - c_i^{(1)}\right) - \left(2 s_i^{(2)} c_i^{(2)} + s_i^{(2)} - c_i^{(2)}\right)}{3} \end{aligned}$$
Thus, the difference between two defuzzified reliability scores is always the sum of the differences between the defuzzified answers to each question, regardless of the particular values of $Score_1$ and $Score_2$. As a result, the metric values are interval-scaled.</td></tr>
</table>

<table>
<tr><th>R3: Quality of the configuration parameters and the determination of the metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The input parameters are the answers to the n questions by experts in the corresponding area. In our CRM mailing campaign scenario, these questions would aim at evaluating the reliability of the customer data with regard to the criteria that are relevant for the campaign (e.g., address, ability-to-pay, willingness-to-pay). Thus, to achieve input parameters of high quality, the answers to these questions need to be gathered by following the standard approaches for questionnaire development and application [Litwin 1995; Marsden and Wright 2010]. Since the remainder of the metric application can be carried out in a formal, automated way, the metric fulfills (R3). This fact is critical to guarantee that the metric values can be used for decision-making, for instance in the CRM mailing campaign scenario.</td></tr>
<tr><th>R4: Sound aggregation of the metric values</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">Yang et al. [2013] do not discuss the application of their metric on different data view levels. Therefore, no aggregation rule is provided. In particular, there is no information regarding the treatment of the n questions in a situation in which the reliability of multiple datasets is assessed (e.g., a possible adaptation of the questions or best practices for consulting experts). In the CRM scenario, this implies that it is not possible to consistently determine the reliability of different databases, for instance, by different external data providers. In that sense, (R4) is not fulfilled.</td></tr>
<tr><th>R5: Economic efficiency of the metric</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The application of the metric requires the answers to each of the n questions by experts as well as the automated determination based on term (3) and the application of a defuzzification operator (4). The last two metric calculations can be done by means of mathematical formulae, both of them in an automated and effective way and at low costs. Moreover, both the survey and calculations are carried out once for the whole dataset and not for each single data value and are also independent of the size of the dataset. Given that in the context of our CRM mailing campaign, the benefits are expected to be significant [Even et al. 2010; Heinrich and Klier 2011], the application of the metric is economically efficient.</td></tr>
</table>

Overall, the metric by Yang et al. [2013] satisfies requirements (R1) to (R3) and (R5), but does not address (R4).

5.4 Metric for Correctness by Hinrichs [2002]

The data quality metric for correctness proposed by Hinrichs [2002] is, on the level of data values, defined as follows:

$$Correctness \coloneqq \frac{1}{d(\omega, \omega_m) + 1} \tag{10}$$

Here, ω is the data value to be assessed, ω_m is the corresponding real-world value and d is a domain-specific distance measure such as, for example, the Euclidean distance or the Hamming distance. A larger difference between ω and ω_m is represented by a larger value of the distance function, which in turn leads to a larger denominator and thus a smaller metric value. The evaluation of the metric with respect to the proposed requirements is presented in Table 6.

Table 6. Evaluation of the Metric by Hinrichs [2002]

<table>
<tr><th>R1: Existence of minimum and maximum metric values</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">If ω perfectly represents the corresponding real-world value ω_m, the distance $d(\omega, \omega_m)$ is determined to be equal to 0 and the metric attains its maximum value of 1. In general, however, the metric values are dependent on the chosen distance function d (which may, for example, be the edit distance, the Euclidean distance or the Hamming distance). This distance function d necessarily varies from dataset to dataset and even between the assessed data values in a particular dataset, as specific distance functions can only be applied to specific data types (e.g., the Euclidean distance function may only be used for numerical data values). Thus, $d(\omega, \omega_m)$ may – dependent on the distance function – become arbitrarily large. Following this, the resulting metric values can indeed be very small while never reaching 0 (as this would require an infinite distance), leading to a violation of (R1). To conclude, the metric does not attain a fixed minimum metric value and (R1) is not fulfilled.</td></tr>
<tr><th>R2: Interval-scaled metric values</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">Common distance measures such as the edit distance, the Euclidean distance or the Hamming distance yield interval-scaled distance values. However, the quotient in the calculation formula inhibits the interval scaling of the resulting metric values: For example, to improve the value of correctness from $\frac{1}{6}$ to $\frac{1}{4}$ (i.e., by $\frac{1}{12}$) , the value of the corresponding distance function has to be decreased from 5 to 3. To improve the value of correctness from $\frac{1}{4}$ further to $\frac{1}{3}$ (i.e., again by $\frac{1}{12}$), only a reduction in distance from 3 to 2 is needed. Thus, the differences of the metric values are in general not meaningful and the metric values are not interval-scaled. Hence, (R2) is not fulfilled.</td></tr>
<tr><th>R3: Quality of the configuration parameters and the determination of the metric values</th><th>(Fulfilled)</th></tr>
<tr><td colspan="2">The metric requires the real-world value corresponding to the data value to be assessed. Determining the real-world value may be resource-intensive in most cases (cf. evaluation of (R5)), but the determination is objective and reliable (as there is exactly one real-world value), and, as long as a well-founded way to determine the value is chosen, valid. For example, in the CRM mailing campaign context, data from external sources (e.g., registration offices or companies such as the German Postal Service, which offer address data) could be used, providing an accurate real-world value. No further configuration parameters are needed, and thus, objectivity, reliability and validity are not violated in this regard. The mathematical formula for calculating the metric values allows for an objective and reliable determination. Finally, the determination of the metric values is valid, because the metric quantifies the data quality dimension correctness according to its definition. Summing up, (R3) is fulfilled.</td></tr>
<tr><th>R4: Sound aggregation of the metric values</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">To determine the metric value at the database level based on its values at the level of relations, Hinrichs [2002] suggests the unweighted arithmetic mean denoted by f in the following. Consider a database D_{l+1} which is decomposed into disjoint relations D_l^h: $D_{l+1} = D_l^1 \cup D_l^2 \cup \ldots \cup D_l^H$ with $D_l^i \cap D_l^j = \emptyset \; \forall i \neq j$ and let further, without loss of generality, the subset D_l^1 be divided into two disjoint subsets $D_l^{1\prime}$ and $D_l^{1\prime\prime}$ at l (i.e., $D_l^1 = D_l^{1\prime} \cup D_l^{1\prime\prime}$, $D_l^{1\prime} \cap D_l^{1\prime\prime} = \emptyset$). Then, let $DQ(D_{l+1}) = f(DQ(D_l^1), \ldots, DQ(D_l^H))$ and $DQ'(D_{l+1}) = f(DQ(D_l^{1'}), DQ(D_l^{1''}), DQ(D_l^2), \ldots, DQ(D_l^H))$. Because f is the unweighted arithmetic mean and the same subsets of D_{l+1} are weighted relatively with $1/H$ or $1/(H+1)$ depending on the particular decomposition used, the equation $DQ'(D_{l+1}) = DQ(D_{l+1})$ does in general not hold, which contradicts a consistent aggregation and thus (R4).</td></tr>
<tr><th>R5: Economic efficiency of the metric</th><th>(Not fulfilled)</th></tr>
<tr><td colspan="2">The metric is based on the comparison of the stored data value and the corresponding real-world value. In many cases, determining the real-world value as input for a data quality metric is (very) resource-intensive as for a large number of data values a real-world comparison is required. For example, in the CRM mailing campaign context, buying external data for a large customer base is (very) expensive and other methods (e.g., trying to contact all customers by phone) similarly require a very high effort. Moreover and in contradiction to an efficient application of the metric, in case the real-world value is known, simply updating the stored data value with the corresponding real-world value would result in perfectly good data quality and the calculation of the metric value would no longer be needed (as this metric value has to represent perfectly good data quality). For example, when the real address of a customer</td></tr>
</table>

is known anyway, applying the metric to measure the correctness of a possibly wrong address provides no additional benefit. Thus, as the metric requires the corresponding real-world values for all stored data values as input, it is not economically efficient and (R5) is not fulfilled.

Overall, the metric by Hinrichs [2002] satisfies (R3), but does not satisfy (R1), (R2), (R4) and (R5).

5.5 Metric for Consistency by Alpar and Winkelsträter [2014]

Alpar and Winkelsträter [2014] define a metric for the consistency of a tuple t as

$$Consistency(t) := \sum_{r \in R} \begin{cases} w^{+}(r), & if\ t\ fulfills\ r \\ w^{-}(r), & if\ t\ violates\ r \\ w^{0}(r), & if\ r\ does\ not\ apply, \end{cases} \tag{11}$$

where R is a set of association rules [Agrawal et al. 1993], $w^{+}(r)$ and $w^{-}(r)$ denote the scoring for a fulfilled and violated association rule, respectively, and $w^{0}(r)$ is the scoring for an inapplicable association rule (which is proposed to be equal to 0). Generally, fulfilled association rules contribute to a higher total score while violated rules lead to a decrease in total score, and tuples with a higher score are assessed as being more consistent. In Table 7, we present the evaluation of the metric based on the requirements.

Table 7. Evaluation of the Metric by Alpar and Winkelsträter [2014]

R1: Existence of minimum and maximum metric values	**(Not fulfilled)**
The metric values depend strongly on the rule set R and the parameters $w^{+}(r)$ and $w^{-}(r)$. The larger the rule set R and the lower the respective weights $w^{-}(r)$, the lower the metric values for tuples violating many rules are. In contrast, the larger the rule set R and the larger the respective weights $w^{+}(r)$, the larger the metric values for tuples fulfilling many rules are. The rule set R and the weights $w^{+}(r)$ and $w^{-}(r)$ necessarily vary from dataset to dataset. As a result, the metric values are neither bounded from below nor from above. Thus, neither a minimum metric value nor a maximum metric value exists and (R1) is not fulfilled.	
R2: Interval-scaled metric values	**(Not fulfilled)**
The parameters $w^{+}(r)$, $w^{-}(r)$ and $w^{0}(r)$ can be set such that the metric value can be interpreted as the percentage of the association rules fulfilled by the tuple. In this case, the metric values are ratio-scaled and hence also interval-scaled. However, the parameters can also represent a non-linear transformation of this setting (e.g., the parameters are a quadratic function), which in turn leads to non-interval-scaled metric values. This is due to the fact that for any two interval scales it is always possible to transform one of them to the other by applying a positive linear transformation of the form $x \mapsto ax + b$ (with $a > 0$) [Allen and Yen 2002]. To conclude, (R2) is not fulfilled. As a consequence, the metric values may lead to wrong evaluations of different decision alternatives. For instance, the difference in consistency of two stored customer addresses is not meaningful and thus cannot be used to determine which customer to select for a CRM campaign.	

R3: Quality of the configuration parameters and the determination of the metric values	**(Fulfilled)**
Association rule mining algorithms [e.g., Agrawal and Srikant 1994] can be used to determine the rule set R in a reliable and objective way. Further, in their application of the metric, the authors propose to use $w^{+}(r) = \mathit{confidence}(r)^{\tau}$, $w^{-}(r) = -\mathit{confidence}(r)^{\tau}$ and $w^{0}(r) = 0$, where $\mathit{confidence}(r)$ represents the confidence of an association rule and $\tau \in \mathbb{N}$ is a calibration parameter. The confidence of an association rule can be calculated reliably and objectively based on simple database queries (e.g., applied to the stored customer data used in the CRM mailing campaign). For τ, the authors suggest a value larger than 25, which is to be verified by experiments. By use of such experiments, τ can then also be determined reliably and objectively. Based upon this, the metric values themselves can be calculated. As the proposed parameters and also the metric itself additionally measure what they should measure and are thus valid, (R3) is fulfilled.	
R4: Sound aggregation of the metric values	**(Not fulfilled)**
Alpar and Winkelsträter [2014] do not provide a definition or an aggregation function to allow the assessment of consistency by means of their metric on a level other than the level of tuples. Hence, it is unclear how to apply the metric and assess consistency on an aggregated level. It follows that (R4) is not fulfilled. Thus, when metric values on an aggregated level are required for decision-making, the metric cannot provide guidance. For instance, the metric cannot be used to assess the consistency of a whole customer database in order to decide whether to perform a data quality improvement measure addressing the database level.	
R5: Economic efficiency of the metric	**(Fulfilled)**
Using $w^{+}(r) = \mathit{confidence}(r)^{\tau}$, $w^{-}(r) = -\mathit{confidence}(r)^{\tau}$ and $w^{0}(r)$, as described in the evaluation of (R3) and suggested for a concrete application, means that the expected costs for applying the metric are low: The rule set R can be determined by a common association rule mining algorithm while the parameters of the metric can be calculated by means of database queries. Similarly, the metric values can be calculated without much effort; all these steps can be performed in an automated and effective way at (rather) low expected costs. The value of the parameter τ needs to be verified by experiments, but this can be done efficiently by preparing a small test set and performing automated tests. In our application context of a CRM mailing campaign, in which significant benefits are to be expected [Even et al. 2010; Heinrich and Klier 2011], the application of the metric is efficient and thus fulfills (R5).	

Overall, while the metric by Alpar and Winkelsträter [2014] fulfills requirements (R3) and (R5), it does not fulfill (R1), (R2), and (R4).

To sum up, the evaluation of the five data quality metrics shows that our requirements are neither trivial nor impossible to fulfill.

6 PRACTICAL IMPLICATIONS

In this section, we discuss the relevance and priority of the requirements with a focus on their practical implications. We provide a combined analysis for (R1) and (R2) as well as separate discussions for (R3), (R4) and (R5). Table 8 summarizes the findings.

R1: Existence of minimum and maximum metric values
R2: Interval-scaled metric values

(R1) and (R2) are of particularly high relevance if, based on the metric values, a decision about different data quality improvement measures or, more generally, about decision alternatives *by means of economic criteria* (cf. economically oriented management of data quality) is made. More precisely: Let us suppose that in a particular application the aim is to just measure the currency of two data values of an attribute and to judge whether the first data value is more up-to-date than the second one (i.e., to make a true/false statement). In this special case, a simple ranking of the metric values for currency of the two data values would be sufficient. Here, one is not interested in the extent of the difference between the metric values for currency of the two data values, nor does one need to know whether the interpretation of one or both metric values for currency suggests (highly) up-to-date or outdated data values.

However, for the large majority of practical applications such a (simple) ranking in the sense of a true/false statement is not sufficient. Rather, based on the metric values, a decision about different decision alternatives *assessed by means of economic criteria* needs to be made. If, in such a case, only a ranking is available, the validation against a specified benchmark (e.g., a required completeness level of 90% of the considered database) is not possible, impeding the use of the metric for decision-making. Furthermore, a ranking cannot support the decision whether the assessed data quality level should be increased based on economic criteria (resp. whether it is even possible to do so). Additionally, when using such a metric, the effects of a data quality improvement measure cannot be clearly compared to its costs. All of these aspects are crucial for an economically oriented management of data quality.

To sum up: A metric might be designed specifically for the context of analyzing the rankings of existing data quality levels or used exclusively in such a context. If this is not the case, but rather a decision about different decision alternatives *assessed by means of economic criteria* (e.g., a comparison of alternative data quality improvement measures) is made based on the metric values, then requirements (R1) and (R2) are highly relevant.

R3: Quality of the configuration parameters and the determination of the metric values

(R3) aims to guarantee that independently of the measuring subjects, one measures what one strives to measure and does so in a correct way. Thus, this requirement covering *validity*, *reliability* and *objectivity* is generally of high importance which can be illustrated by the example of assessing the data quality dimension currency. In practical applications, *internal validity* is of particular relevance. Internal validity first addresses that the underlying definition of currency ("object of interest") is indeed measured by the metric. Second, it also ensures that significant changes in the metric values (i.e., the dependent variable) are indeed caused by a change in the variables which influence currency and not by extraneous factors (control variables). In contrast, *external validity* is primarily only of high relevance if the metric is just applied to a sample of the dataset, but the results are used to derive statements regarding the whole dataset. *Reliability* aims to guarantee that the metric leads to equal or very similar results (i.e., a high stability of the results) in repeated assessments of the same data (e.g., in the course of time) and to thus ensure a correct measurement in this regard. *Objectivity* is particularly relevant to allow both an automated data quality assessment and obtaining metric values which are independent of external influences (e.g., different interviewers).

Overall, data quality metrics not fulfilling (R3) can provide insufficient metric values (cf. above). Regarding an economically oriented management of data quality, this is, for instance, problematic when evaluating the data quality level before and after conducting a data quality improvement measure. A metric not fulfilling (R3) cannot deliver trustworthy results with respect to the actual change in the data quality level. Thus, a data quality improvement measure may be evaluated as effective but may not actually improve data quality at all or only by a very small margin.

To sum up, when designing and applying a metric, the following points need to be considered:

(a) It is important to analyze which data values, metadata and parameter values are required to instantiate and apply a data quality metric: If extensive historical data (either from internal or external sources; big / open data) is available, the required data values and parameters (in particular, the configuration parameters) can be determined in a valid, objective and reliable way using statistical techniques. If such a data basis is not available, for instance expert estimations are needed, which also have to be obtained in a transparent and verifiable way.

(b) Where possible, metrics should be formally defined such that – as long as the required data values and parameters are clearly defined – the calculation rule ensures (R3) (in particular, objectivity and reliability). If the calculation rule cannot be formally defined, the calculation of the metric values needs to be described in a stepwise, transparent way and as clear as possible to allow an intersubjective application. In any case, the correspondence between what is to be measured (in particular, an exact definition of the respective data quality dimension) and what is actually measured (operationalization of the defined data quality dimension) needs to be ensured.

R4: Sound aggregation of the metric values

(R4) is of high relevance if the assessment or the selection of decision alternatives is not just based on the isolated data quality assessment of a single data value. More precisely: Let us consider an application in which the aim is just to measure the completeness of the data values of an attribute, independently from each other. Let further the individual metric values be directly used for decision-making, for example such that in case no data value (or a data value semantically equivalent to 'NULL') is stored, apply action *a*; otherwise do not apply any action. In this case, an isolated decision based on the level of data values is performed, which does not require any aggregation. However, decisions in practice, for example regarding the application of data quality improvement measures, are usually not just based on a single data value or individual data values considered in an isolated way. Rather, this requirement is of particular practical relevance in many decision situations that rely on the data quality of (large) sets of data values. For example, the data quality of a larger part of a customer database (or even the whole database) may be considered to decide whether to conduct a marketing campaign.

To sum up: If a metric was not explicitly designed for statements regarding the data quality of single data values (resp. it is not only used in such situations), but rather is designed or used to express the data quality of multiple data values in a single metric value, (R4) is particularly relevant. The higher the importance of this aggregated metric value for decision support, the higher the relevance of (R4).

R5: Economic efficiency of the metric

(R5) is of particularly high priority if assessing data quality by means of a metric results in substantial costs resp. the metric values are used for a decision with potentially large costs and benefits. Especially against the background that in practice low data quality often results in high costs [Experian Information Solutions 2016; IBM Big Data and Analytics Hub 2016; Forbes Insights 2017], this requirement needs to be taken into account already in the design of a metric. More precisely: Let us consider an application in which the aim is just to measure the completeness of the data values of a single attribute in a relation with around 100 tuples. The assessment is conducted manually by a single person within a time span of five minutes (i.e., the costs for determining both the configuration parameters and the metric values are negligible). This person stores the result of the assessment (i.e., the percentage of complete data values according to the metric) just for documentation purposes in a file, no further analysis is conducted and no decisions at all are based on the result of the assessment (i.e., any additional payoff from the application of the metric is irrelevant). In such cases, in practice, evaluating the efficiency of the metric – in particular in comparison to alternative metrics which might possibly allow a slightly faster counting – is hardly necessary. Similarly, one might argue that evaluating the efficiency of metrics is not required in the application case of a data quality assessment mandatory due to legal regulations (e.g., in risk management). Here, one could reason analogously that the evaluation of the efficiency is not relevant for the decision whether to apply a metric. However, this argument may fall short: Even in the case of a mandatory assessment, a company may again evaluate the economic efficiency of two or more possible metrics to select the most appropriate one. Thus, in many cases, (R5) is highly relevant from a practical perspective. Moreover, (R5) is also of particular importance when assessing data quality as part of a data governance or data quality management initiative, as these are generally aimed at economic efficiency.

To sum up: Data quality metrics are usually not designed for assessing data quality in cases of low economic relevance of the assessment (when both the additional payoffs as well as the costs resulting from an application of the metric are negligible). Thus, the relevance of (R5) is obvious. This relevance increases the higher the expected costs resp. the expected benefits from both measuring data quality and the decisions based on the assessment are. Table 8 summarizes the findings with regard to decision situations for which specific requirements are of particular relevance.

Table 8. Practical Situations with particular Relevance for specific Requirements

	(R1) and (R2)	(R3)	(R4)	(R5)
Of particular relevance in practical situations in which ...	- ... decision alternatives are assessed by means of economic criteria. - ... multiple, related data quality assessments are performed, for instance over time. - ... a particular focus resides on the interpretability of the metric values. - ... improvement measures and their impacts on data quality are compared or evaluated.	- ... both configuration parameters and metric values are not absolutely trivial to determine (in contrast to situations where, for instance the configuration parameters are given or can be obtained effortlessly). - ... multiple, related data quality assessments are performed, for instance over time.	- ... metric values on different aggregation levels are relevant for decision-making. - ... the decision relies on the data quality of (large) sets of data values. - ... multiple, aggregations are performed and the results are compared. - ... the aggregation of metric values is necessary for the determination of one metric value on an aggregated level.	- ... potentially large costs and benefits emerge. - ... an efficient metric has to be selected amongst different feasible metrics. - ... multiple, related data quality assessments are performed, for instance over time.

7 CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

In this paper, we propose a set of five requirements for data quality metrics to support both decision-making under uncertainty and an economically oriented management of data quality. Our requirements contribute to existing literature in two ways. First, as opposed to existing approaches, which are fragmented and leave room for interpretation, we present a set of clearly defined requirements, thus making it possible to easily and transparently verify them. This is very important for practical applications. Second, in contrast to existing works, we justify our requirements based on a sound decision-oriented framework. If such a framework is missing, it is neither possible to substantiate the relevance of the requirements nor is it clear what happens if a requirement is not met. As a result, our requirements are essential for the evaluation of existing metrics as well as for the design of new metrics (e.g., in the context of Design Science Research). Based on our requirements, inadequate metrics, which may lead to wrong decisions and economic losses, can be identified and improved. The applicability and efficacy of the proposed requirements are demonstrated by means of five well-known data quality metrics. The application to the metric for completeness by Blake and Mangiameli [2011] reveals the existence of metrics which satisfy all requirements. The application to the metrics by Ballou et al. [1998], Yang et al. [2013], Hinrichs [2002] and Alpar and Winkelsträter [2014], however, shows that the requirements are not trivial to fulfill. Both results are crucial from a methodical and practical point of view.

The proposed requirements constitute a first but essential step to support both decision-making under uncertainty and an economically oriented management of data quality. Nevertheless, they also have limitations. First, they are designed for data quality metrics concerning data views and therefore do, for instance, not directly consider data quality metrics addressing the quality of data schemes. However, in future research, the ideas underlying the derivation of the requirements can be transferred analogously to other types of data quality metrics. Moreover, as already discussed for many other sets of requirements (e.g., in the context of software engineering), it is not possible to prove the completeness and sufficiency of a set of requirements. Indeed, extending a set of requirements is an iterative process, which should consider both theoretical and practical aspects. Thus, future research should extend the proposed set of requirements in a well-founded manner.

REFERENCES

AGRAWAL, R., IMIELIŃSKI, T., AND SWAMI, A. 1993. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, P. BUNEMAN AND S. JAJODIA, Eds. ACM Press, New York, NY, 207–216.

AGRAWAL, R., AND SRIKANT, R. 1994. Fast Algorithms for Mining Association Rules. In *Proceedings of the 20th International Conference on Very Large Data Bases*, J. B. BOCCA, M. JARKE AND C. ZANIOLO, Eds. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 487–499.

ALLEN, M., AND CERVO, D. 2015. *Multi-Domain Master Data Management. Advanced MDM and Data Governance in Practice*. Morgan Kaufmann.

ALLEN, M.J., AND YEN, W.M. 2002. *Introduction to measurement theory*. Waveland Press, Long Grove Ill.

ALPAR, P., AND WINKELSTRÄTER, S. 2014. Assessment of data quality in accounting data with association rules. *Expert Systems with Applications 41*, 5, 2259–2268.

AZUMA, M. 2001. SQuaRE: the next generation of the ISO/IEC 9126 and 14598 international standards series on software product quality. In *ESCOM (European Software Control and Metrics conference)*, 337–346.

BALLOU, D., WANG, R., PAZER, H., AND TAYI, G.K. 1998. Modeling information manufacturing systems to determine information product quality. *Manag Sci 44*, 4, 462–484.

BATINI, C., AND SCANNAPIECO, M. 2006. *Data quality: concepts, methodologies and techniques*. Springer, New York.

BATINI, C., AND SCANNAPIECO, M. 2016. Data Quality Dimensions. In *Data and Information Quality*. Springer, 21–51.

BLAKE, R., AND MANGIAMELI, P. 2011. The effects and interactions of data quality and problem complexity on classification. *Journal of Data and Information Quality (JDIQ) 2*, 2, 8.

BRIAND, L.C., MORASCA, S., AND BASILI, V.R. 1996. Property-based software engineering measurement. *IEEE Trans Software Eng 22*, 1, 68–86.

BUHL, H.U., RÖGLINGER, M., MOSER, F., AND HEIDEMANN, J. 2013. Big data. A fashionable topic with(out) sustainable relevance for research and practice? *BISE 5*, 2, 65–69.

BUREAU INTERNATIONAL DES POIDS ET MESURES. 2006. *The international system of units (SI)*. National Institute of Standards and Technology, Paris.

CAI, L., AND ZHU, Y. 2015. The challenges of data quality and data quality assessment in the big data era. *Data Science Journal 14*.

CAI, Y., AND ZIAD, M. 2003. Evaluating completeness of an information product. In *AMCIS (2003)*, 2273–2281.

CAMPANELLA, J. 1999. *Principles of Quality Costs: Principles, Implementation and Use*. ASQ Quality Press, Milwaukee.

CAPPIELLO, C., AND COMUZZI, M. 2009. A utility-based model to define the optimal data quality level in IT service offerings. In *ECIS (2009)*.

CAPPIELLO, C., DI NOIA, T., MARCU, B.A., AND MATERA, M. 2016. A quality model for linked data exploration. In *International Conference on Web Engineering*, 397–404.

COZBY, P., AND BATES, S. 2012. *Methods in behavioral research*. McGraw-Hill Higher Education, New York.

DEBATTISTA, J., AUER, S., AND LANGE, C. 2016. Luzzu—A Methodology and Framework for Linked Data Quality Assessment. *Journal of Data and Information Quality (JDIQ) 8*, 1, 4.

DRIANKOV, D., HELLENDOORN, H., AND REINFRANK, M. 1996. *An Introduction to Fuzzy Control*. Springer Berlin Heidelberg; Imprint; Springer, Berlin, Heidelberg.

EPPLER, M.J. 2003. *Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes*. Springer, Berlin.

EVEN, A., AND SHANKARANARAYANAN, G. 2007. Utility-driven assessment of data quality. *Database Adv Inform Syst 38*, 2, 75–93.

EVEN, A., SHANKARANARAYANAN, G., AND BERGER, P.D. 2010. Evaluating a model for cost-effective data quality management in a real-world CRM setting. *Decision Support Systems 50*, 1, 152–163.

EXPERIAN INFORMATION SOLUTIONS. 2016. *Building a business case for data quality*. https://www.edq.com/globalassets/white-papers/building-a-business-case-for-data-quality-report.pdf. Accessed 19 July 2017.

FAN, W. 2015. Data Quality. From Theory to Practice. *SIGMOD Rec 44*, 3, 7–18.

FEIGENBAUM, A.V. 2004. *Total quality control*. McGraw-Hill Professional New York.

FISHER, C.W., CHENGALUR-SMITH, I., AND BALLOU, D.P. 2003. The impact of experience and time on the use of data quality information in decision making. *Inform Syst Res 14*, 2, 170–188.

FISHER, C.W., LAURIA, E.J.M., AND MATHEUS, C.C. 2009. An accuracy metric: Percentages, randomness, and probabilities. *Journal of Data and Information Quality (JDIQ) 1*, 3, 16.

FLOOD, M., JAGADISH, H.V., AND RASCHID, L. 2016. Big data challenges and opportunities in financial stability monitoring. *Banque de France, Financial Stability Review 20*.

FORBES INSIGHTS. 2017. *The Data Differentiator. How Improving Data Quality Improves Business*.

HEINRICH, B., AND HRISTOVA, D. 2014. A Fuzzy Metric for Currency in the Context of Big Data. In *ECIS (2014)*.

HEINRICH, B., AND HRISTOVA, D. 2016. A quantitative approach for modelling the influence of currency of information on decision-making under uncertainty. *Journal of Decision Systems 25*, 1, 16–41.

HEINRICH, B., KAISER, M., AND KLIER, M. 2007. How to measure data quality? A metric-based approach. In *ICIS (2007)*.

HEINRICH, B., AND KLIER, M. 2011. Assessing data currency-a probabilistic approach. *Journal of Information Science 37*, 1, 86–100.

HEINRICH, B., AND KLIER, M. 2015. Metric-based data quality assessment—Developing and evaluating a probability-based currency metric. *Decision Support Systems 72*, 82–96.

HEINRICH, B., KLIER, M., AND GÖRZ, Q. 2012. Data quality assessment: a metric-based approach to quantify the currency of data in information systems. *Z Betriebswirtsch 82*, 11, 1193–1228 (in German).

HEINRICH, B., KLIER, M., AND KAISER, M. 2009. A procedure to develop metrics for currency and its application in CRM. *Journal of Data and Information Quality (JDIQ) 1*, 1, 1.

HINRICHS, H. 2002. *Datenqualitätsmanagement in Data-Warehouse-Systemen. Dissertation*. Universität Oldenburg.

HIPP, J., GÜNTZER, U., AND GRIMMER, U. 2001. Data Quality Mining-Making a Virtue of Necessity. In *6TH ACM SIGMOD DMKD (2001)*, 52–57.

HIPP, J., MÜLLER, M., HOHENDORFF, J., AND NAUMANN, F. 2007. Rule-Based Measurement Of Data Quality In Nominal Data. In *ICIQ (2007)*, 364–378.

HÜNER, K.M. 2011. *Führungssysteme und ausgewählte Maßnahmen zur Steuerung von Konzerndatenqualität*. Dissertation. Universität St. Gallen.

HÜNER, K.M., SCHIERNING, A., OTTO, B., AND ÖSTERLE, H. 2011. Product data quality in supply chains: the case of Beiersdorf. *Electronic Markets 21*, 2, 141–154.

IBM BIG DATA AND ANALYTICS HUB. 2016. *Extracting business value from the 4 V's of big data*. http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data. Accessed 19 July 2017.

IBM GLOBAL BUSINESS SERVICES. 2012. *Analytics: Big Data in der Praxis*. IBM Global Business Services, Armonk.

ISO/IEC 25020. 2007. *Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Measurement reference model and guide 35.080*.

JIANG, Z., SARKAR, S., DE, P., AND DEY, D. 2007. A framework for reconciling attribute values from multiple data sources. *Manag Sci 53*, 12, 1946–1963.

JONES, B.D. 1999. Bounded rationality. *Annu Rev Polit Sci 2*, 1, 297–321.

KHATRI, V., AND BROWN, C.V. 2010. Designing data governance. *Communications of the ACM 53*, 1, 148–152.

KPMG. 2016. *Now or never - 2016 Global CEO Outlook*. Accessed 31 July 2017.

LAUX, H. 2007. *Decision theory*. Springer Gabler, Wiesbaden (in German).

LEE, Y.W., STRONG, D.M., KAHN, B.K., AND WANG, R.Y. 2002. AIMQ: a methodology for information quality assessment. *Inform Manag 40*, 2, 133–146.

LEVY, Y., AND ELLIS, T.J. 2006. A systems approach to conduct an effective literature review in support of information systems research. *Imforming Science 9*, 1, 181–212.

LI, F., NASTIC, S., AND DUSTDAR, S. 2012. Data Quality Observation in Pervasive Environments. In *CSE (2012)*, 602–609.

LITWIN, M.S., Ed. 1995. *How to Measure Survey Reliability and Validity*. The Survey Kit 7. Sage, Thousand Oaks, Calif.

LOSHIN, D. 2010. *The practitioner's guide to data quality improvement*. Morgan Kaufmann.

LUKOIANOVA, T., AND RUBIN, V.L. 2014. Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances in Classification Research Online 24*, 1, 4–15.

MARSDEN, P.V., AND WRIGHT, J.D., Eds. 2010. *Handbook of survey research*. Emerald, Bingley.

MOORE, S. 2017. *How to Create a Business Case for Data Quality Improvement*. http://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/. Accessed 19 July 2017.

MOSLEY, M., BRACKETT, M., AND EARLEY, S., Eds. 2009. *The DAMA guide to the data management body of knowledge enterprise server version*. Technics Publications, LLC, Westfield.

NITZSCH, R. von. 2006. *Entscheidungslehre*. Verlag Mainz, Mainz.

ORR, K. 1998. Data quality and systems theory. *Communications of the ACM 41*, 2, 66–71.

OTTO, B. 2011. Data Governance. *BISE 3*, 4, 241–244.

PARSSIAN, A., SARKAR, S., AND JACOB, V.S. 2004. Assessing data quality for information products: impact of selection, projection, and Cartesian product. *Manag Sci 50*, 7, 967–982.

PETERSON, M. 2009. *An introduction to decision theory*. Cambridge University Press, Cambridge.

PIPINO, L.L., LEE, Y.W., AND WANG, R.Y. 2002. Data quality assessment. *Communications of the ACM 45*, 4, 211–218.

REDMAN, T.C. 1996. *Data quality for the information age*. Artech House, Boston.

SARSFIELD, S. 2009. *The data governance imperative*. IT Governance Publishing.

SAS INSTITUTE. 2013. *2013 Big data survey research brief*. SAS Institute, Cary, N.C.

SIMON, H.A. 1956. Rational choice and the structure of the environment. *Psychological review 63*, 2, 129–138.

SIMON, H.A. 1969. *The sciences of the artificial*. MIT press, Cambridge.

STEVENS, S.S. 1946. On the theory of scales of measurement. *Science 103*, 2684, 677–680.

TALEB, I., EL KASSABI, H.T., SERHANI, M.A., DSSOULI, R., AND BOUHADDIOUI, C. 2016. Big Data Quality. A Quality Dimensions Evaluation. In *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences*, 759–765.

WANG, R.Y. 1998. A product perspective on total data quality management. *Communications of the ACM 41*, 2, 58–65.

WANG, R.Y., STOREY, V.C., AND FIRTH, C.P. 1995. A framework for analysis of data quality research. *IEEE Trans Knowl Data Eng 7*, 4, 623–640.

WEBER, K., OTTO, B., AND ÖSTERLE, H. 2009. One size does not fit all--a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ) 1*, 1, 4.

WEBSTER, J., AND WATSON, R.T. 2002. Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly 26*, 2, 13–23.

WECHSLER, A., AND EVEN, A. 2012. Using a Markov-Chain Model for Assessing Accuracy Degradation and Developing Data Maintenance Policies. In *AMCIS (2012)*.

YANG, L., NEAGU, D., CRONIN, M.T.D., HEWITT, M., ENOCH, S.J., MADDEN, J.C., AND PRZYBYLAK, K. 2013. Towards a Fuzzy Expert System on Toxicological Data Quality Assessment. *Molecular informatics 32*, 1, 65–78.

ZIKMUND, W., BABIN, B., CARR, J., AND GRIFFIN, M. 2012. *Business research methods*. Cengage Learning, Mason.

APPENDIX A: NOTATION

Table 9. Notation

Notation	Definition
s_j	State of nature, $j \in \{1, \dots, n\}$
$S = (s_1, s_2, \dots, s_n)$	Vector of all considered states of nature
$w(s_j)$	Probability of occurrence for a state of nature s_j
a_i	Decision alternative, $i \in \{1, \dots, m\}$
$A = (a_1, a_2, \dots, a_m)$	Vector of all considered decision alternatives
p_{ij}	Payoff if alternative a_i is chosen and state of nature s_j occurs
$P_i = (p_{i1}, p_{i2}, \dots, p_{in})$	Vector of the payoffs for alternative a_i and all considered states of nature
$E(a_i, P_i, S)$	Expected payoff without considering data quality for alternative a_i, given a vector S of states of nature and a vector P_i of payoffs for alternative a_i
DQ $DQ(\dots)$	Data quality metric value of the considered data value or set of data values
$E(a_i, DQ, P_i, S)$,	Expected payoff when considering data quality for alternative a_i, given a vector S of states of nature, a vector P_i of payoffs for alternative a_i, and a value of the data quality metric DQ
M	Supremum/maximum of the considered metric values
l	Data view level with $l \in \{1, \dots, L\}$
D_l	A dataset at data view level l
D_l^h	A subset of the dataset D_l, $h \in \{1, \dots, H\}$
$\widetilde{D}_l^k$	A subset of the dataset D_l, $k \in \{1, \dots, K\}$
f	Aggregation function

APPENDIX B: REQUIREMENTS FOR DATA QUALITY METRICS PROPOSED BY HÜNER ET AL. [2011]

Table 10. Requirements for Data Quality Metrics proposed by Hüner et al. [2011]

Requirement	Description of the proposed Requirement
Cost/benefit	The costs for the definition and the calculation of the data quality metric values ought to be in a positive ratio (< 1) to the benefits (controlled error potential).
Definition of measurement frequency	The instants of time at which the values of a data quality metric are calculated should be defined.
Definition of measurement point	The measurement point (e.g., data repository, process, department) of a data quality metric should be defined.
Definition of measurement procedure	The instrument (e.g., survey, software) to determine the data quality metric value should be defined.
Definition of scale	A scale (e.g., percentage, school grades, time) should be defined for a data quality metric value.
Limitation of the application data	For a data quality metric, the data to be applied to (e.g., material master, European customers) should be defined.
Escalation process	For a data quality metric appropriate measures should be defined depending on certain threshold values (i.e., metric values to initiate data quality measures).
Validity range	A range should be defined for a data quality metric in which its values are valid.
SMART criteria	A data quality metric should fulfill the SMART criteria (specific, measurable, attainable, relevant and time-bounded).
Disturbance variables	The metadata of a data quality metric should contain information about possible disturbance variables (i.e., it should describe possible events or impacts which may distort the values of the data quality metric).
Responsibility	For a data quality metric clear responsibilities should be defined such as to whom and which values of the data quality metric are reported, who is responsible for the maintenance of the metric (e.g., up-to-date/meaningful definition, implementation of the measurement procedure).
Comparability	A data quality metric should be defined so that its values can be compared to those of other metrics (data quality metrics or process metrics).
Comprehensibility	For a data quality metric metadata should be available, which describes its purpose and the correct interpretation of its values.
Use in SLAs	It should be possible to use data quality metric values in Service Level Agreements.
Visualization	It should be possible to visualize the values of a data quality metric (e.g., time series, diagrams).
Repeatability	It should be possible to determine the values of a data quality metric not only once, but multiple times.
Target value	For a data quality metric a target value should be defined.
Assignment to a data quality dimension	It should be possible to assign a data quality metric to one or more data quality dimensions.
Assignment to a business problem	It should be possible to assign a data quality metric to a specific (company-specific) business problem.
Assignment to a process figure	It should be possible to assign a data quality metric to one or more process figures.
Assignment to the company strategy	It should be possible to assign a data quality metric to one or more strategic goals of the company.