DOES THE EU INSURANCE MEDIATION DIRECTIVE HELP TO IMPROVE DATA QUALITY? – A METRIC-BASED ANALYSIS

Heinrich, Bernd

Kaiser, Marcus

Klier, Mathias

Abstract

Most financial service companies view the requirements imposed by the insurance mediation directive (IMD) as a tiresome administrative burden, causing high costs but providing little benefit. In contrast, this article argues that the documentation requirements demanded by the IMD lead to a higher data quality and also to higher economic benefits. A significant improvement is proclaimed regarding the data quality dimensions correctness and completeness. To substantiate this hypothesis, we develop two metrics based on existing approaches; one for correctness and one for completeness. The development is guided by six general requirements on data quality metrics. Moreover, a case study analyses the influence of the IMD on data quality by applying these metrics in a case study: Based on data from a major German insurance company, we illustrate how economic benefits arise from documenting particular categories of customer data (e.g. the customer's professional background, his financial circumstances and his goals), while the documentation of other customer data categories does not provide similar benefits.

Keywords: Data Quality Metrics, Insurance Mediation Directive, Financial Services.

1 INTRODUCTION

Latest changes in the European commission's regulatory instruments, as e.g. the insurance mediation directive (IMD, (European Parliament and the Council 2002)), intend to improve consumer protection. The directive includes a particular measure to protect the customer more effectively: the documentation, which requires the financial services consultant to capture the customer data upon which the advice is based. Many financial services providers (FSP) consider the IMD as an administrative burden, resulting in high costs without providing the corresponding added value. In contrast to this position, we argue that the documentation requirements lead to a higher data quality (DQ). As research shows, DQ has a significant impact on companies' profitability (Heinrich & Kaiser & Klier 2007, Parssian & Sarkar & Jacob 2004). This also holds during an advisory process for financial services, as poor DQ – e.g. incorrect or incomplete data – will often result in wrong recommendations, thus decreasing both, consultant's success and customer satisfaction.

Before quantifying the economic consequences of DQ, DQ itself has to be quantified. Therefore, we discuss how metrics for selected DQ dimensions can be designed based on two objectives: (1) Enabling the measurement of DQ, and (2) Analysing the economic consequences of the documentation requirements by the IMD. The paper is structured as follows: Section 2 states general requirements the DQ metrics have to meet (Note: these requirements must not be confused with the documentation requirements imposed by the IMD). In section 3, we design formal metrics for the DQ dimensions correctness and completeness. Section 4 presents an application of the metrics in practice. Subsection 4.1 describes the proposed documentation requirements. In subsection 4.2, a case study evaluates the designed metrics by applying them in a project at a major German insurance company:

The goal was to determine the correlation between the quality of customer data of an insurance branch and its sales. Thereby the economic consequences of the documentation requirements can be analysed. Section 5 concludes the paper and critically reflects the results.

2 REQUIREMENTS ON DATA QUALITY METRICS

This paper aims at quantifying DQ by means of metrics for particular dimensions. The identification and classification of DQ dimensions is treated from both a scientific and a practical point of view by many publications (e.g. (Eppler 2003, Lee et al. 2002, Wang & Storey & Firth 1995)). Following Redman (1996), we can for instance distinguish the dimensions listed in Table 1.

Category	DQ Dimensions		
Conceptual View	• Content • Scope	• Level of Detail • Composition	• View Consistency • Reaction to Change
Data Values	• Accuracy/Correctness • Completeness	• Consistency • Currency	
Data Representation	• Appropriate Format • Interpretability	• Format Flexibility • Portability	• Efficient Use of Storage • Ability to represent NULL values

Table 1. DQ categories and dimensions.

Juran (1998) refers to the quality of *Data Values* as Quality of Conformance and to the *Conceptual View* as Quality of Design. Quality of Conformance represents the degree of correspondence between the specification and the existing realisation in the IS (e.g. data schemes vs. set of stored customer data). In contrast, Quality of Design denotes the degree of correspondence between the users' expectations and the specification of the information system (IS) (e.g. specified by means of data schemes). Whereas Quality of Conformance refers to the – more objective – measurement of the correspondence between the specified data schemes and the existing data values, quantifying Quality of Design is (mostly) subjective, as it measures the correspondence between the users' expectations and the specified data schemes. Therefore, we do not deal with the *Conceptual View* and leave aside *Data Representation* (as a more technical perspective on DQ) as well. Instead we focus on Quality of Conformance, as the quality of the values currently stored in the IS is of high relevance with respect to the documentation requirements imposed by the IMD. From the four dimensions concerning the Quality of Conformance, we select correctness and completeness as these dimensions are particularly important within the financial advisory process: Missing information for instance prohibits to analyse whether a (new) product is suitable for a customer, while incorrect data may lead to wrong advice causing huge damage for both, the customer and the FSP. In addition, for reasons of liability the consultant is forced to ensure the data used for recommendations are correct.

As DQ measures are frequently developed on an ad hoc basis to solve specific, practical problems (Pipino & Lee & Wang 2002) they often contain a high degree of subjectivity (Cappiello & Francalanci & Pernici 2004). To ensure a scientific foundation and allow for an evaluation of the metrics, we state the following general *requirements on DQ metrics*. They have proven particularly useful for an economic-oriented DQ management (cf. (Even & Shankaranarayanan 2007, Heinrich & Kaiser & Klier 2007)):

R 1. *[Normalisation]* An adequate normalisation is required to allow value comparison of the metrics (e.g. to compare different levels of DQ over time). In this context, DQ metrics are typically ratios with a value between 0 (poor) and 1 (perfect).

R 2. *[Interval scale]* To support both the monitoring of the changes of the DQ level over time and the economic evaluation of measures, we require the metrics to be interval scaled.

R 3. *[Interpretability]* The quantification has to be "easy to interpret". Hence, the DQ metrics have to be comprehensible.

R 4. *[Aggregation]* DQ has to be quantified on different levels. For a relational data model this implies the ability to quantify DQ on the level of attribute values, tupels, relations/views as well as on the whole database so that the values have consistent semantic interpretation on each level. In addition, the metrics must allow aggregation of quantified results to the subsequent higher level.

R 5. *[Adaptivity]* To quantify DQ in a goal-oriented way, the metrics have to be adaptable to the context of a particular application.

R 6. *[Feasibility]* To ensure applicability, the metrics have to be based on measurable input parameters. When defining metrics, measurement methods must be defined and where exact measurement is not possible or too cost-intensive, alternative (rigorous) methods (e.g. statistical) have to be proposed.

3 DESIGN OF THE METRICS

In the following section, we design DQ metrics for correctness and completeness. To meet the requirement of aggregation (R 4) the metrics are constructed "bottom up". I.e., a metric on level $n+1$ (e.g. completeness on the level of tupels) is based on the corresponding metric on level n (e.g. completeness on the level of attribute values).

3.1 Correctness

In concordance with Batini and Scannapieco (2006) (and their definition of accuracy) we define correctness as the closeness between a value w_I and a value w_R, considered as the correct representation of the real world phenomenon that w_I aims to represent. The approach by Hinrichs (2002) is – to the best of our knowledge – the only formally noted approach which (1) develops a metric especially for correctness, (2) mainly focuses on Quality of Conformance, and (3) aims at an objective, goal-oriented measurement. However, Hinrichs' approach comes along with several problems (Heinrich & Kaiser & Klier 2007) which are mainly due to two reasons: Firstly, he uses distance functions normalised to the interval $[0; \infty]$ leading to weaknesses related to normalisation (cf. R 1) and interpretability (cf. R 3). Secondly, his functional equation is based on a quotient, which results in values not being interval scaled (cf. R 2) and both absolute and relative changes of the metric can not be interpreted.

Therefore, we develop a new metric: Let w_I be an attribute value and w_R the corresponding attribute value in the real world. $d(w_I,w_R)$ is a domain-specific distance function quantifying the closeness between w_I and w_R. To assure the metric being normalised to the interval [0; 1] (cf. R 1), we use – in contrast to Hinrichs – distance functions normalised to the interval [0; 1]. Examples for such distance functions are: $d_1(w_I, w_R) := \begin{cases} 0 & \text{if } w_I = w_R \\ 1 & \text{else} \end{cases}$, which is independent of the field of application, $d_2(w_I, w_R) := \left(\frac{|w_I - w_R|}{\max\{|w_I|, |w_R|\}} \right)^{\alpha}$ with $\alpha \in R^+$ for numeric, metrically scaled attributes and n-grams, edit (or Levenshtein) distance or Hamming distance for strings (all normalised to the interval [0; 1]). Based on such distance functions, we avoid a quotient and define the metric on the level of attribute values as follows:

(1) $$Q_{Corr.}(w_I, w_R) := 1 - d(w_I, w_R)$$

An example demonstrates how the metric works: Prior to a mailing campaign, the correctness of the attributes "income" and "year of birth" shall be evaluated, as the offer (about an old age pension scheme) is addressed at people belonging to a certain age group and income class. Since both attributes are numeric, we use the distance function $d_2(w_I,w_R)$. To assure the adaptivity according to

(R 5) on the level of attribute values, the parameter α has to be chosen. This can for instance be done by experts' estimations or a sensitivity analysis. The smaller α, the higher the penalty for smaller deviations and vice versa: The relevance of selecting an appropriate value for α can be shown in our example: In case of the attribute "year of birth" even small deviations shall be penalised as a deviation of only 1% (e.g. the years of birth w_I=1957 and w_R=1977) has a huge impact: When for example offering an old age pension scheme, 50 year old customers have to be treated differently than 30 year old clients. Therefore, the distance function must sensitively react even to small deviations. Thus we require α<1 (leading to an over-proportional increase of the function for $\frac{|w_I - w_R|}{\max\{|w_I|, |w_R|\}} << 1$). For example choosing α=0.01 – in case of a deviation of 1% – results in a value of only 4.5% for the metric of correctness. Thus relatively small deviations would result in a very low value of the metric. In contrast, when considering the correctness of the attribute "income", smaller deviations might not be as critical. This is due to the fact that an offer which is intended for a particular income could also be interesting for customers whose income is close to the intended one. Since the distance function ought to tolerate smaller deviations, we set α>1 (resulting in an under-proportional increase of the function if $\frac{|w_I - w_R|}{\max\{|w_I|, |w_R|\}} << 1$). For α=1.10 and a deviation of 1% (e.g. the stored annual income is 80,000€ in contrast to the real income of 80,800€) we receive a value of 99.5% for the correctness of the attribute "income". After determining α, both the value of the distance function and the DQ metric in formula (1) has to be computed. We avoided a quotient and decided to use the functional term (cf. (1)) to ensure the interpretability of the resulting values. Therefore, the metric is interpretable referring to the related distance function (cf. R3) while at the same time a value range of [0;1] is ensured (cf. R 1). Moreover, the values of the metric are interval scaled (cf. R 2). This ensures – in combination with meeting requirement (R 3) – that the metric is applicable within an economic management of DQ. Constructing the metric "bottom up", we have to define it on the levels of tupels, relations/views and database next (due to length restrictions, we omit the formulas here and refer to subsection 3.2).

Comparing the attribute values stored in the IS to their real world counterparts is crucial in order to quantify correctness. However it is important to check the attribute values for their syntactical representation before this comparison. Otherwise, the distance function would return distance values greater than 0 (and the metric for correctness less than 1), although there is no semantic difference. Consider the common abbreviation "Rd" for "Road", for example. If the abbreviation is not replaced by the long version (or vice versa), the distance function and the metric will behave as follows (if we use the edit distance, Levenshtein distance or Hamming distance):

d("Abbey Rd","Abbey Road") > 0 $\Rightarrow$ $Q_{Corr.}$("Abbey Rd","Abbey Road") < 1

Moreover, the metric for correctness should only be applied to attribute values which *semantically differ* from *NULL* (i.e. attribute values which are complete, cf. subsection 3.2). Otherwise, the results of the metric will be distorted if we quantify the correctness of attribute values with a value of the metric for *completeness* of 0 (Note: The purpose of the metric for correctness is to quantify the difference between stored attribute values and their existing real world counterparts). This shall be demonstrated by the following example: If we replace *NULL* values with arbitrary values - as practiced by some companies - we can manipulate the value of the metric for correctness: Assume w_I=*NULL*, w_R=8 and the dummy variable for replacing *NULL* values w_D=0. Since the distance function is 1 for *NULL* values, replacing *NULL* with 0 might result in a distance value lower than 1. As a consequence, the value for correctness would be higher than 0, thereby distorting the value of the metric for correctness:

$1 \geq d(0,8) \Rightarrow Q_{Corr.}(0,8) \geq 0$

Applying the metric for correctness only to attribute values whose results for completeness are positive, the metric value can be computed automatically, if the real world counterpart of each considered attribute value is known. This can only be achieved by conducting a survey which is

expensive for huge sets of data. Hence we propose to consider only a sample of attribute values that is representative for the whole data set and survey the real world counterpart of these attribute values. Thereby, conclusions can be drawn from the sample for the whole data set by means of statistical procedures (cf. (Helfert 2002)). Thus, we get an unbiased estimator for $Q_{Corr.}$. E.g., it is possible to buy up-to-date address data from external sources for a sample of the customer base. These bought addresses can be compared to the address data stored in the IS. The value of the metric $Q_{Corr.}$ can then be used as an unbiased estimator for the correctness of the whole address data set.

3.2 Completeness

Due to our focus on Quality of Conformance, completeness means that stored data attributes must have values that semantically differ from *NULL*. In this context, *NULL* is not a defined value, but a mere wildcard for unknown or non-present values. We assume that there are no "mismembers" within a relation, (tupels which are part of this relation, but should not be). That is why the approach by Parssian, Sarkar and Jacob (2004) is not suitable for our purposes. Furthermore, we act under the "closed world assumption" as defined by Batini and Scannapieco (2006).

A number of promising approaches to quantify completeness exist in literature (Even & Shankaranarayanan 2005, Hinrichs 2002, Naumann & Freytag & Leser 2004), others are summed up in Scannapieco and Batini (2004). When designing our metric for completeness in the following, we refer to, adopt or precise them where necessary.

Let w be an attribute value stored in the IS. Then the metric for completeness on the level of attribute values $Q_{Compl.}(w)$ is defined as follows:

$$(2) \qquad Q_{Compl.}(w) := \begin{cases} 0 & \text{if } w = NULL \text{ or } w \text{ is semantically equal to } NULL \\ 1 & \text{else} \end{cases}$$

To meet R 1, the quality of an attribute value considering completeness is valuated as 0 (minimum) if the attribute has no entry or the entry is a (default) value semantically equal to *NULL* (e.g. a dummy value). Otherwise, the result of the metric is 1 (maximum). Thereby, the results on this level are – not surprising – clearly interpretable (cf. R 3), too.

Problems arise when the reason for a *NULL* entry is not unavailable data, but the fact that the corresponding value does not exist in the real world. An example is "spouse's name" in case of singles. Then the entry *NULL* does not represent incompleteness and the value of the metric should be 1 and not 0. Such problems have to be solved in order to provide an adequate documentation of customer data as required by the IMD, since it makes a huge difference with respect to the needs of a customer whether there is a spouse and his/her name is unknown or whether the person is single. A feasible solution is the introduction of indicators telling whether or not the corresponding value exists in the real world. For instance, the attribute "spouse's name" can be assigned the value "not married", if the marital status is acquired as "single". Thus, the corresponding attribute in the database is assigned a value and the metric returns the (intended) value of 1. Prior to applying the metric, the attributes have to be analysed and emerging problems have to be addressed. After this verification of all attribute values, the metric can be applied to the whole data set automatically.

Based on the level of attribute values, we define the metric on the level of tupels (cf. R 4). Here, T is a tupel with the values $T.A_1, \ldots, T.A_{|A|}$ for the attributes $A_1, \ldots, A_{|A|}$ and $g_i \in [0;1]$ is the relative importance of A_i regarding completeness. Thus we define the metric for completeness on the level of tupels based on the metric on the level of attribute values as a weighted arithmetic mean:

$$(3) \qquad Q_{Compl.}(T) = \frac{\sum_{i=1}^{|A|} Q_{Compl.}(T.A_i) g_i}{\sum_{i=1}^{|A|} g_i}$$

The results are normalised (cf. R 1), interval scaled (cf. R 2) and interpretable (cf. R 3): E.g. a difference of 0.2 between two tupels in the same context always indicates that the tupel with the higher value of the metric contains 20% more attribute values than the tupel with the lower value.

Using the formula above, the quantification of the completeness of a tupel bases on the quantification of the completeness of those attribute values which are part of the tupel (cf. R 4). In contrast, the next level in Naumann, Freytag and Leser (2004) are not tupels, but attributes, i.e. a column of a relation (or view). This approach seems less applicable for our purposes, because our (business) focus is on the customer who is normally represented by a tupel. Quantifying the completeness of all values in a column is not helpful for managing the contact to a particular customer more effectively. Therefore, we do not take into account the approach by Naumann, Freytag and Leser (2004) in the following.

As formula (3) demonstrates, we use the weights g_i to enable a flexible and goal-oriented adaptation of the metrics (R 5). This is reasonable, as the importance of an attribute might differ depending on the particular context. For instance, during an advisory session attributes like "year of birth" or "income" are especially important, whereas "last name" and "address" are less relevant in this situation. If a FSP wants to improve DQ in order to support his consultants more effectively, higher weights ought to be put on "year of birth" and "income" than on "last name" or "address". By contrast, if the data are to be used in a forthcoming mailing, it is most important that "last name" and "address" are available. Therefore, the weights on "last name" and "address" have to be higher than the ones on "year of birth" and "income". The option of assigning different weighting to attributes is an advantage compared to existing approaches (e.g. (Scannapieco & Batini 2004)), as they allow for more flexibility – if needed – regarding different application settings.

Next, let R be a non empty relation or a view. Then the completeness of R bases on the arithmetic mean of the completeness of the tupels T_j in R (j = 1, 2, …, $|T|$) and is – to meet (R 1) to (R 3) – defined as:

$$(4) \qquad Q_{Compl.}(R) = \frac{\sum_{j=1}^{|T|} Q_{Compl.}(T_j)}{|T|}$$

Using this formula, all values on the level of tupels are weighted equally and summed up on the level of relation. I.e. all tupels being part of the relation are equally important. In contrast, the formula for quantifying completeness on the level of relations in Even and Shankaranarayanan (2005) includes weights in order to emphasise particular tupels. This may be reasonable for instance in cases when the customer lifetime value (CLV) is stored within the data base and the value of this attribute may affect the value of the metric for completeness (i.e. storing the data of high valued customers is more important than of low valued customers). However, this can also have negative effects: E.g. when solely the CLV of some customers increases while no data are added, the value of the metric for completeness of the relation will increase, as well. Moreover, if the weights change over time, the metric – not surprisingly – returns different values for each quantification even if all other factors and data remain the same. As we want to avoid such effects (in order to avoid manipulations), we omitted weighting factors when quantifying the completeness of relations.

Finally, let D be a data set (e.g. a database) which can be represented as a disjoint decomposition of the relations or views R_k (k = 1, 2, …, $|R|$). I.e. the whole data set can be decomposed into pairwise non-overlapping relations R_k, so that each attribute in the data set is assigned to exactly one of the relations, or formally noted: $D=R_1 \cup R_2 \cup \ldots \cup R_{|R|}$ and $R_i \cap R_j = \varnothing \ \forall i \neq j$. (Note: in cases when a key attribute is part of several relations or views, it has to be weighted with a positive value only once. This avoids a multiple consideration within the metric for completeness and does not prohibit the applicability of the metric). Hence, we define the completeness of a data set D (based on the completeness of the relations R_k (k=1, 2, …, $|R|$)) as follows:

(5) $$Q_{Compl.}(D,R) := \frac{\sum_{k=1}^{|R|} Q_{Compl.}(R_k) g_k}{\sum_{k=1}^{|R|} g_k}$$

Hinrichs (2002) defines the completeness of a data set *D* via an unweighted arithmetic mean. Thereby, relations that are very important for the given context influence the value as strongly as relations being not important at all. In contrast, the weights $g_k \in [0;1]$ used in our definition allow incorporating the relative importance of particular relations according to the specific objective of the quantification. Moreover, according to Hinrichs the resulting value depends on the disjoint decomposition of the database into relations. This makes it difficult to evaluate the completeness of a database objectively. For example, Hinrichs relatively weights a relation R_k for k≠2 with 1/|R| when using the disjoint decomposition $\{R_1, R_2, R_3, \ldots, R_{|R|}\}$, whereas the same relation is only weighted with 1/(1+|R|) when using the disjoint decomposition $\{R_1, R_2', R_2'', R_3 \ldots, R_{|R|}\}$ for $R_2' \cup R_2'' = R_2$ and $R_2' \cap R_2'' = \emptyset$.

When the completeness of a database is to be quantified from a neutral, more technical point of view (i.e. without a specific context), the weights have to be chosen according to the size of the particular relations (taking also into account the number of tupels and attributes). Let $|T|_k$ be the number of tupels and $|A|_k$ the number of attributes of tupels within a relation R_k. Then the weights of all attributes have to be chosen equally and the weight of a relation R_k has to be determined via $g_k = \frac{|A|_k \cdot |T|_k}{\sum_{m=1}^{|R|} \left(|A|_m \cdot |T|_m \right)}$ for all $k=1, \ldots, |R|$. However, in many cases the metric will be applied in a specific context. Then, the weighting factors should be chosen in reconciliation with the operating department (R 5).

Quantifying the completeness of a data set can be done by applying the metric above via corresponding SQL statements. Thereby it can easily be determined whether or not an attribute is assigned a value (cf. R 6). Furthermore, this procedure can be applied to the whole data set in an automated way. Only *NULL* values which are supposed to have no value cause a problem, which can be solved by means of indicators as described above.

4 APPLICATION OF THE METRICS

After designing the metrics for correctness and completeness, subsection 4.1 describes the documentation requirements. Then, subsection 4.2 studies the consequences of the documentation requirements in depth by presenting how the designed metrics were applied in a project at a major German insurance company.

4.1 The Insurance Mediation Directive

The Directive 2002/92/EC of the European Parliament and the European Council on insurance mediation aims at ensuring a higher level of consumer protection by imposing professional reliability to be verified by a registration authority. Furthermore, it specifies information and documentation requirements for insurance intermediaries (which we refer to as “consultants” in the following). As we intend to analyse the effects of the IMD on DQ, we will focus on the documentation requirements: “Prior to the conclusion of any specific contract, the insurance intermediary shall at least specify, in particular on the basis of information provided by the customer, the demands and the needs of that customer as well as the underlying reasons for any advice given to the customer on a given insurance product […] All information to be provided to customers […] shall be communicated […] on paper or on any other durable medium available and accessible to the customer” (European Parliament and the Council 2002). The documented information has to be provided to the customer in a clear and comprehensible textual form prior to signing a contract. This instruction given by the law is quite vague. In many cases the “demands and needs” as well as the “underlying reasons for any advice” will

base on personal data of the customer. Consider the following example: A customer wants to buy a household insurance. A frequently used rule says that the sum insured should be calculated according to the formula: *living space (of the property) in sqm x 650 €*. According to the IMD, this rule and the parameters (i.e. the living space of the customer's property) must be documented, since they are the "underlying reasons" for calculating the sum insured which reflects the customer's need. In case of more complex products (as e.g. life insurances for old age provision), still more pieces of information about the customer must be asked in order to justify the advice. Due to the variety of products, data concerning many spheres of life have to be documented. Besides master data, e.g. data about other banking or insurance products the customer possesses (even with other FSPs) should be acquired in order to determine whether the customer has further needs. In addition, data about the professional or family background are important for products concerning provision.

4.2 Case study

In this subsection, we describe the application of the metrics within a business environment. Furthermore, the case study shall substantiate the thesis that the documentation requirements by the IMD cause a better DQ – quantified by means of the metrics – and result in higher sales. For reasons of confidentiality, the figures had to be changed and made anonymous. Nevertheless, the procedure and the basic results remain the same.

The metrics were used to quantify DQ within a project in collaboration with a major German insurance company which offers not only insurance contracts, but also private investments. The insurance company engaged the authors to analyse the DQ within their branches. Moreover, propositions for the forthcoming implementation of the IMD should be made. Due to space restrictions, we can only illustrate the application of one metric. Since the completeness of customer data is directly affected by the IMD, we decided for this metric.

Within the project the DQ of 75 branches were analysed by means of the metric. We deliberately selected branches with very similar structures, i.e. high resemblance regarding clientele, number of employees, usage of advertising material and other characteristics (e.g. urban vs. rural area, sold product categories). The 75 branches represented about 38% of all 197 branches (population) with a structure as described above. The objective was to examine the quality of the existing customer data. This could be done easily, since all branches used the same sales IS provided by the insurance company. It allowed to store a huge amount of customer and sales data, based on a given relational database scheme. We also assured that the branches did not document additional data to a noteworthy extent in other systems or paper-based. The data attributes stored in the IS were categorised into seven groups. In the following, a few representative attributes for each group are given in brackets:

1. Customer master data (name, sex, address, date of birth)
2. Contract data (ongoing/expired/cancelled contracts, reasons for signing/cancelling a contract)
3. Contact history (date and place of the consultation, topics of the consultation, offered contracts, reasons for configuration of the offered contracts, customer's feedback concerning the offers)
4. Customer's professional background (pursued profession, professional status, current employer, income, learned profession, former employers)
5. Customer's family background (marital status, data concerning partner, data concerning children, relatedness to other customers of the branch)
6. Customer's financial circumstances (investment in different asset classes, illiquid/liquid assets)
7. Customer's goals (professional and private goals, (future) needs for financing and insurance including amount and point of time, need for old age provision)

The existing data were analysed via the metric for completeness based on the data specification of the sales IS. First, an analysis tool was installed and a dump of the database was created. After that, all *NULL* values which do not stand for a missing value, but are due to the fact that no corresponding real

world value exists (as e.g. in case of “spouse’s name” for customers with the marital status “single”, cf. above) were replaced by adequate indicators. Thirdly, the tool read in the customer data from the dump by means of corresponding SQL statements and computed the value of the metrics for completeness as noted above. Moreover, it provided further functionality for detailed analysis. Some of this was used for the results described in the following.

The analysis revealed that the values of the metric for data assigned to categories 1 to 3 was higher than for the one of data within the categories 4 to 7. However, apart from category 1, the branches differed (significantly) with respect to the completeness of the data they documented. This is remarkable, since the branches were urged to systematically document customer data far before the project. Let us for example consider the data assigned to category 7: 11 branches achieved a value of the metric for completeness within the interval]0.5;0.6], i.e. between 50% and 60% of the specified attributes were filled for the customers of these branches. In contrast, the value was in the interval [0;0.1] for 17 branches.

The insurance company wanted to know – concerning the application of the IMD – whether there is a correlation between the completeness of the data and the success of a branch regarding its sales. To answer this question the insurance company provided sales figures of the analysed branches. The sales were measured by means of the goal realisation level, expressed by the ratio between the realised sales of a branch and the target sales. The latter were set equally for all branches of the population by the insurance company. To examine the correlation between completeness of data and sales, firstly we assigned each branch according to its value of the metric for completeness to one of the intervals [0;0.1],]0.1;0.2], …,]0.9;1]. This was done for each of the seven categories separately. Then the average goal realisation level of the branches assigned to an interval was calculated. Figure 2 depicts the results for category 7, which we choose as an illustration example, since the data assigned to it are most directly affected by the documentation requirements of the IMD (note that no branch achieved a higher value of the metric for completeness than 0.6).

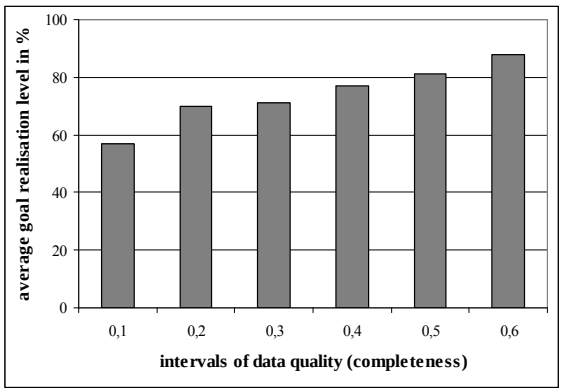

Figure 2. Average goal realisation level depending on the metric for completeness (category 7)

The results of Figure 2 indicate that there is a positive correlation between the value of the metric for completeness of customer data and the success of a branch (Note that the results do not mean that the success solely depends on the completeness of the attributes of category 7). The correlation coefficient for the given example is 0.81. One has to keep in mind that the time spent for documenting is not available for other sales activities. This fact can not be neglected, because each branch had approx. 800 customers in average whose data have to be documented. Unfortunately, precise data concerning the time spent on documenting were not available. However, knowing that the branches were similarly structured and assuming that those branches with a low completeness used the time saved by not documenting for other sales activities, it seems that documenting the data of category 7 comes along with a corresponding economic benefit: Although documentation takes time which can not be spent for other sales activities, those branches which document more data are also more successful.

Taking a closer look, these results can be justified since the data in category 7 mainly relates to future financial services needs of the customer. Such information helps the consultant to pro-actively offer products which are adequate for the customer's needs. This augments the probability of selling a product to the customer, thereby increasing the sales figures. As described above, the IMD especially requires documenting the customer's financial desires and needs. Therefore, even those branches with a low value of the metric for completeness for their customer data are obliged to store these pieces of information in the future (at least when giving an advice to a customer). Considering the results above we can conclude, that this aspect of the IMD is not only an administrative burden which solely causes costs, but might go along with higher sales.

Data assigned to most of the other categories (except for category 1) are not as directly affected by the IMD as category 7. Nevertheless, we also analysed the completeness of data in categories 1 to 6 and its correlation with the success of a branch. The branches were divided into three groups according to their goal realisation level: 26 "successful" branches achieved a goal realisation level higher than 80%, whereas the goal realisation level of 23 "non-successful" branches was lower than 65%. The other 26 branches had a goal realisation level between 65% and 80% and their value of the metric for completeness was for all categories between the ones of the "non-successful" and the "successful" branches (for illustrative purposes, we leave them aside in the following). Figure 3 depicts the results:

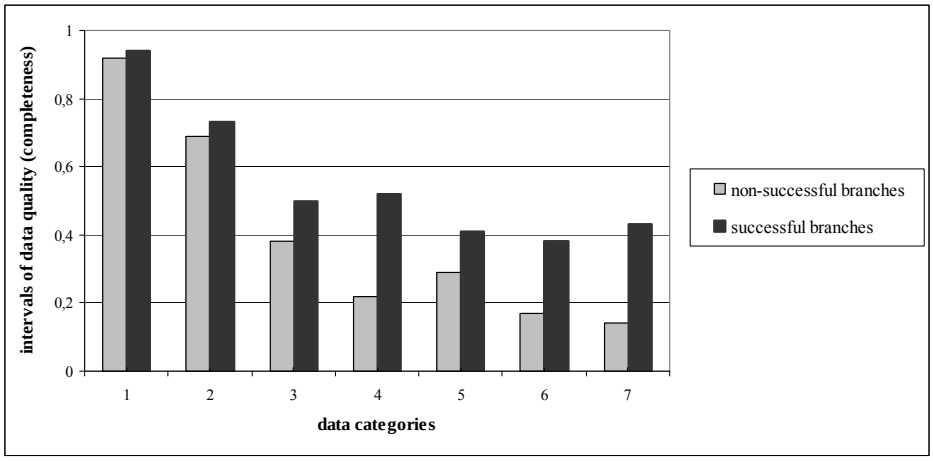

Figure 3. Values of the metric for completeness of customer data within different categories

The results for categories 1 and 2 are not surprising: These are mainly customer master data and contract data and therefore crucial for daily business. The difference regarding the completeness of customer data between "successful" and "non-successful" branches is quite small. Therefore, completeness of categories 1 and 2 does not seem to be sufficient to explain the difference between "successful" and "non-successful" branches. Next, we can state noticeable differences between the completeness of customer data assigned to categories 3 and 5: In average, the completeness for "successful" branches is about 35% higher than for "non-successful" branches. These results indicate that the completeness of contact history data and of data about the customer's family background is notably higher for "successful" than for "non successful" branches. The highest difference – besides category 7 – for the completeness of customer data between "successful" and "non-successful" branches can be recognised for the categories 4 and 6. In both cases, the average metric value for completeness for "successful" branches is more than 130% higher than for "non-successful" ones, nearly reaching the value of category 7. A basic result of the analysis is that the customer data of "successful" branches are more complete than the data of "non-successful" ones. Moreover, there are three types of data which can be classified according to (1) documentation requirements by the IMD and (2) positive correlation with success:

1. Data which have to be documented according to the IMD (1) and the quality of the data is positively correlated with success (2): For this kind of data (e.g. data of category 7), the IMD

provides benefit to all consultants by requiring the documentation, since the results revealed that "successful" branches reached a higher value of the metric for completeness than "non-successful" branches.

2. Data which have to be documented according to the IMD (1) and the quality of the data is not correlated with success (2): E.g. the IMD requires the master data (category 1) to be documented; however the difference between "successful" and "non-successful" branches concerning the completeness of this kind of data is low. The documentation of such data is – besides being necessary for managing the customer contact in parts – indeed an administrative burden.
3. Data which do not have to be documented according to the IMD (1) and the quality of the data is positively correlated with success (2): For instance, the data concerning the customer's professional background (category 4) is not influenced directly by the IMD but "successful" branches document this kind of data much more completely than "non-successful" branches.

These results of the case study illustrate that the quality of customer data is positively correlated with the success of an insurance branch (the results for correctness were similar to the ones for completeness). However, it becomes also obvious that not all customer data are positively correlated with success. Hence, in preparing for the IMD it is important to closely examine for which data categories and attributes a high data quality is valuable. For some attributes it might provide economic benefit to document more than is required by the IMD. Concerning the project, it might be criticised that the actual use of the data by the consultants (e.g. during an advisory process or within a mailing campaign) was not analysed. I.e. (as mentioned above) that other factors might also be correlated with the success of a branch.

After the analysis, the insurance company defined a documentation guideline for the IMD: The branches were urged to completely (and correctly) as possible document particular data categories and data attributes, which were identified based on the results of the project. These data categories and data attributes were partly beyond the ones required by the IMD, e.g. the attributes in category 4. Since the documentation guidelines took effect recently, an ex post analysis is not yet available. However, first interviews with the consultants lead to positive feedback.

The insurance company quantified its DQ for the first time and gained the insights described above by applying the metrics to their customer data. After seeing the results and convinced by the automated (and therefore comparatively inexpensive) way of quantifying DQ, the management decided to apply the metrics to other data as well, e.g. to transaction and product data.

5 CONCLUSION

The paper analysed how the DQ dimensions correctness and completeness can be quantified in a goal-oriented and economic manner. Therefore, we designed metrics for these two dimensions which meet – in contrast to existing approaches – the general requirements, like adaptivity and feasibility, stated at the beginning of the design process. The metrics were applied in cooperation with a major German insurance company within a project for preparing its documentation guidelines and its IS for the IMD. They turned out to be useful for the given purpose as they allow quantifying the impact of incomplete and incorrect data in an automated way, delivering interpretable results for economic analyses.

As mentioned above, it might be criticised that the application of the metric for correctness can not be entirely automated, as the comparison of all attribute values to their real world counterparts is very time-consuming and therefore costly. The authors are - based on the experiences during the project - currently working on the adaptation of existing statistic procedures which shall help to lower the costs of this comparison. Furthermore, as also described, non-existing values for an attribute might be a problem concerning the metric for completeness.

Both metrics can also be used in other domains, where completeness is important and data might be wrongly acquired, e.g. in the area of customer relationship management: For instance it does not seem

reasonable to use an attribute value as a selection criterion for a campaign, which has an estimated value of 40% of the metric for correctness over all customers. Such decisions can be supported by the metrics designed in this article. Their application is also reasonable if a data set is processed, provided that the assumptions above hold. Therefore, we encourage further research in order to apply the metrics in management, production or logistic processes.

References

Batini, C. and Scannapieco M. (2006). Data Quality. Concepts, Methodologies and Techniques (Data-Centric Systems and Applications). 1st Edition. Springer, Berlin.

Cappiello, C., Francalanci C. and Pernici B. (2004). Data quality assessment from the user's perspective. In Proceedings of the IQIS 2004, International Workshop on Information Quality in Information Systems (Naumann, F. and Scannapieco, M., Eds), 68-73, ACM, New York.

Eppler, M.J. (2003). Managing information quality. 1st Edition. Springer, Berlin.

European Parliament and the Council (2002). Directive 2002/92/EC on insurance mediation. Accessed: 2007/10/24, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0092:EN:NOT

Even, A. and Shankaranarayanan G. (2005). Value-driven data quality assessment. In Proceedings of the 10th International Conference on Information Quality (Naumann, F., Gertz, M. and Madnick, S.E., Eds), 221-236, MIT Press, Cambridge/Boston.

Even, A. and Shankaranarayanan, G. (2007). Utility-Driven Assessment of Data Quality. The DATA BASE for Advances in Information Systems, 38 (2), 75-93.

Heinrich, B., Kaiser M. and Klier M. (2007). How to measure data quality? – a metric based approach. In Proceedings of the 28th International Conference on Information Systems (ICIS) (Rivard, S. and Webster, J., Eds), Association for Information Systems, Montreal.

Helfert, M. (2002). Proaktives Datenqualitätsmanagement in Data-Warehouse-Systemen - Qualitätsplanung und Qualitätslenkung. 1st Edition. Logos, Berlin.

Hinrichs, H. (2002). Datenqualitätsmanagement in Data Warehouse-Systemen. Oldenburg.

Juran, J.M. (1998). How to think about Quality. Juran's Quality Handbook (Juran, J.M. and Godfrey, A.B., Eds.), McGraw-Hill, New York, 2.1-2.18.

Lee, Y.W., Strong, D.M., Kahn, B.K. and Wang, R.Y. (2002). AIMQ: a methodology for information quality assessment. Information & Management, 40 (2), 133-146.

Naumann, F., Freytag, J. and Leser, U. (2004). Completeness of Integrated Information Sources. Information Systems, 29 (7), 583-615.

Parssian, A., Sarkar, S. and Jacob, V.S. (2004). Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product. Management Science, 50 (7), 967-982.

Pipino, 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. 1st Edition. Artech House, Boston.

Scannapieco, M. and Batini C. (2004). Completeness in the relational model: a comprehensive framework. In Proceedings of the 9th International Conference on Information Quality (Chengalur-Smith, I.N., Raschid, L., Long, J. and Seko, C., Eds), 333-345, MIT Press, Cambridge/Boston.

Wang, R.Y., Storey, V.C. and Firth, C.P. (1995). A Framework for analysis of data quality research. IEEE Transactions on Knowledge and Data Engineering, 7 (4), 623-640.