ANALYZING DATA QUALITY INVESTMENTS IN CRM: A MODEL-BASED APPROACH

(Completed Research)

Markus Helfert
School of Computing, Dublin City University
markus.helfert@computing.dcu.ie

Bernd Heinrich
Department of Information Systems, University of Augsburg
bernd.heinrich@wiwi.uni-augsburg.de

Abstract Known as Customer Relationship Management (CRM) in recent years a concept for business that focuses customers is often discussed in research and practice. After initially extraordinary expectations, numerous CRM projects fail. Thereby as one of the major reasons an overestimated and poor data quality is very frequently mentioned. Many authors assume a positive correlation between data quality and CRM, but nevertheless this can often not be (obviously) justified. This article aims to contribute to this research and analyses data quality investments in customer relationship management. By providing an explanation model we analyze the interdependences. With this model, it can be shown that data quality investments do not necessarily result in lasting or intensive customer relationships. In addition to the scientific contribution provided by the model, it builds the basis for derive recommendations for practice.

Key Words: Data Quality, Data Quality Effects, Customer Relationships, Customer Relationship Management

1. INTRODUCTION

In order to explain the success but also the failure of projects in customer relationship management aspects of data quality are frequently mentioned [e.g. 1; 25]. But however, it is still an unanswered question how data quality affects customer relationships. Regarding its importance, as it is often stated in research and practice, there is still only little research addressing this question. As a basis for addressing this research problem, an explanation model to study and represent interdependences between CRM and data quality would be essential. This article presents a theoretical contribution to this research problem and provides a foundation for further research.

The article is structured as following. First a reflection of related research takes place, by which we describe the research problem and characterize its context. This provides not only a definition for data quality and customer relationships but also a model for representing its dependencies. Having developed such a model, it can be applied for analyzing data quality investments in CRM. In conclusion, we review the developed model and discuss future research directions.

2. RELATED WORK

Various publications mention interdependences between data quality and customer relationship management [e.g. 1; 7; 8; 9; 23; 25; 28], and frequently (mostly dogmatic) a positive correlation between high data quality and customer satisfaction is implied. For example, most authors implicit presume that high data quality enables product individualization and thus consequently strengthen the relationship to its customers. But, most authors so far do not discuss in deep how data quality influences customers' purchase decision or how data quality affects the relationship intensity. These statements would be fundamental, in order to analyze the effectiveness of measures taken to improve data quality. For the lack of research two reasons may be considered:

(1) First an essential foundation for research in form of a commonly accepted and formal data quality definition is still not present. The term 'data quality' (or often synonymously used 'information quality') is examined in numerous publications [e.g. see literature in 18; 33], resulting in a multiplicity of descriptions, definitions, criteria lists and classification frameworks for different application areas [see for some application examples e.g. 3; 20; 21; 22; 24]. In these publications a user and application-oriented data quality view is most dominant, whereas data quality is determined regarding its fitness for use [e.g. 21]. By a set of (application context specific) quality criteria, like for example correctness, completeness, consistency, relevance and timeliness as well as interpretability, availability, data accessibility and data security most literature concretized data quality further [see e.g. 33]. Although these literatures provide already a general basis, no systematic in data quality criteria, definitions and dependencies can be found. In particular no formalized data quality definition exists so far. However this would be necessary in order to study the interdependences between data quality and customer relationships.

(2) Besides data quality we have to analyze what a relationship-oriented interaction in contrast or in addition to an usual product-oriented or transaction-oriented interaction is. In order to be able to point out the potential effects of data quality, it is necessary to study and (formally) represent the construct of customer relationship. In literature exist a variety of definitions and conceptions. Numerous authors refer that a relationship is to be understood as a sequence of reciprocally connected, not coincidentally realized transactions. Therefore it is rather a holistic, continuously interaction with so-called episodes than single purchases, which can be unambiguously and clearly separated from each other [15; 27]. But what is the essence of the 'internal connection' and what relevance must this connection have in order to speak about a relationship?

Numerous, partially also different opinions exist about it. Many papers focus, like e. g. [12], that „a series of transactions gradually transforms into a relationship as a result of the social exchange between buyer and seller. A relationship is thus something much more than a series of transactions, and contains dimensions of power, cooperation, commitment and trust to name but a few." In contrast to this other authors emphasize the long-term, economic objectives of the partners as well as the character of an investment [5], which are lost as sunk costs if the relationship is terminated. Other papers name also barriers of exit in the sense of different costs, like search costs and learning costs or risk factors as characteristic for a relationship, whereby its longevity is clearly rejected as a necessary criterion [e.g. 31]. Similar to this short discussion a number of further sources can be found in literature, which point out (partially contradictory) criteria and cases, in which a relationship could or does exist respectively does not exist [e.g. 27]. To that extent O'Malley and Tynan summarize: „Despite more than ten years of academic and practitioner interest in this area, understanding of the nature of business to consumer relationships has advanced little. […] Given the diversity in operational approaches employed, and the lack of accepted definitions, it has become impossible to delimit the domain. The boundaries are completely permeable and elastic." [26].

The discussion above shows that a suitable definition for data quality or customer relationship does not exist or cannot be adopted. Thus it is very important to provide for further work a theory-driven, conceptual basis for the terms data quality and customer relationship in the next two chapters.

3. FORMALIZING DATA QUALITY

As discussed, a more formalized data quality definition is needed and therefore in the following data quality will be clarified from the author's point of view. According to general quality definitions [6; 14], quality of data can be differentiated into *Quality of Design* and *Quality of Conformance* [19; 30; 32]. Quality of design refers to the degree of correspondence between user requirements and their concretion in specifications. In contrast quality of conformance enfolds the degree of correspondence between specifications and production processes and its products. Transferring this concept to information systems, in following a formal definition of data quality is provided.

In general an explicit or implicit specification describes requirements for information system components, like for example software programs and functions and data models. In following such a specification is called I_t^{spec}, which describes a specification for an information system at time t. A data request of a data user u at time t shall be called $I_{t,u}^{demand}$. Because in general not all requirements can be included into I_t^{spec}, I_t^{spec} does usually not entirely conform to $I_{t,u}^{demand}$ from all users (e.g. an attribute value is requested, which is not considered in the data model). Data provided by the information system at time t shall be called I_t^{supply} (e.g. the provided customer data). Again, due to insufficient implementation of I_t^{spec} and real world constrains, I_t^{supply} does in general not entirely conform to I_t^{spec} (e.g. some attribute values are incomplete or incorrect).

Based on this, quality of design and quality of conformance can be defined. First, a (standardized) quality function of data user u at time t is used to describe the quality of design as $Q_{t,u}^{design}(I_t^{spec}, I_{t,u}^{demand}) \rightarrow [0;1]$, whereby the value 0 represents no quality and the value 1 represents maximum quality. Second, a (standardized) quality function $Q_t^{conform}(I_t^{spec}, I_t^{supply}) \rightarrow [0;1]$ describes the quality of conformance between specification and data provided. This function is independent from the data user, whereby the value 0 represents no quality and the value 1 represents maximum quality. In general, it can be assumed that increasing I_t^{spec} results in higher Q_t^{design} and that increasing $I_{t,u}^{demand}$ results in lower Q_t^{design} (exceptions have to be considered later). Similar applies to quality of conformance $Q_t^{conform}$, whereby increasing I_t^{spec} results in lower $Q_t^{conform}$ and increasing I_t^{supply} results in higher $Q_t^{conform}$.

Having formalized the two elements of data quality, data quality management objective function is to maximize the total quality Q_t^{total} over all application areas [18], which can be described with the optimization variables I_t^{spec}, I_t^{supply} and $I_{t,u}^{demand}$ as (whereby weighting and efficiency remain so far unconsidered):

$$Q_t^{total} = \sum_u Q_{t,u}^{design}(I_t^{spec}, I_{t,u}^{demand}) + Q_t^{conform}(I_t^{spec}, I_t^{supply}) \Rightarrow \max!$$

In other words, data quality managements aims to consolidated best possible the various user requirements into a specification and fulfill the specification best possible by the information system.

In general data users and their tasks determine the data demand $I_{t,u}^{demand}$. In this article we do not consider influencing the data demand and thus we assume that it is predetermined. Assuming predetermined data demand, direct data quality improvements can be done

(a) by an optimization of the specification I_t^{spec} or

(b) by an (qualitative) increasing of the data provided I_t^{supply}.

Case (a) refers to questions regarding information requirement analysis and its specifications, which shall also not be subject of this article. In contrast case (b) includes measures for increasing the quality of conformance, whereas in principle two options exist:

1. Increasing the data supply I_t^{supply} can take place with measures of using and incorporating new data (e.g. completion of data by means of new customer data). In the following these measures are represented by the variable $D_{SUP} \in [0;1]$, whereby $D_{SUP} = 0$ means that none of the requested data are present. On the other hand $D_{SUP} = 1$ describes that all necessary data are present. If we do not consider additional external data (e.g. purchasing of additional customer data), the data supply does in particular depend on business transactions, since these are the basis for data gathering. Further can be assumed that first transactions result in the largest increase in data (e.g. address data, customer's basic requirements etc.). Additional transactions, in particular if they are identical to previous, will result in smaller increasing of data (e.g. transaction data for representing the customer contact history). Thus, D_{SUP} has in relation to accomplished business transactions a decreasing marginal utility.
2. A qualitative increase in the sense of improving the data correctness of I_t^{supply} can also take place via measures of data cleansing, which is considered as part of reactive data quality management [18]. In addition to data cleansing a qualitative increase of I_t^{supply} is possible by measures of process improvement in the sense of an improvement of completeness and correctness (e.g. modification of data gathering processes or data transfer processes). These measures are assigned to proactive data quality management [18]. In further, reactive and proactive measures are represented by the variable $D_{QM} \in [0;1]$, whereby $D_{QM} = 0$ means that neither data cleansing nor high-quality processes are accomplished; respectively $D_{QM} = 1$ means that the measures are at its maximum.

In summary, assuming that I_t^{spec} is given, the quality of conformance can be described as

$$Q_t^{conform}\left(I_t^{spec}, I_t^{supply}(D_{SUP}, D_{QM})\right) \rightarrow [0,1] \qquad (1)$$

4. DEFINING RELATIONSHIPS FROM A CUSTOMER VIEW

As discussed, because of the lack of precise definitions in literature, it is necessary to define first the construct of customer relationship in the context of our research. We follow a customer's perspective [e.g. 10], because nowadays customers, in particular valuable customers, select business relations independently. This is to the fact, that in general in saturated markets companies cannot (autonomously) decide, with which customers they would like to establish a business relationship (at least not in terms of intensive customer relationships). In contrast to purely transaction-oriented interaction, in the following we clarify relationship-oriented interactions and develop a model. This model is then extended to analyze data quality investments and is based on the following assumptions:

(P1) A rational-acting customer has a utility preference function under certainty. This means he or she can assign a real utility value $\Phi(a)$ to each offer $a_i \in A$ submitted by a provider using a mapping $\Phi: A \rightarrow \Re$. Thereby different alternatives can be prioritized in relation to its value. Thus an alternative a_i for committing a *business transaction* is in relation to another alternative a_k [superior/inferior/equivalent] if the utility value $\Phi(a_i)$ is [>/</=] to $\Phi(a_k)$.

(P2) Within a relevant time period, the customer's needs are homogeneous according to the considered product (e.g. several homogeneous needs for insurance policies or petrol). By using the utility preference function it is possible to rate not only the core product or transaction but also additional services like e.g. kindness of employees or the provider's image.

(P3) A decision model of one period with t subperiods is considered. This means, that in particular providers and customers make decisions at the beginning of the time period without any time preferences. The decisions are then implemented within the subperiods.

(P4) A continuous model is assumed. In particular, data quality measures can create any necessary value for the variables D_{SUP} and D_{QM} and we do not distinguish between measures for reactive and proactive management.

(P5) The entire transaction volume, which results from the satisfaction of customer needs, is arbitrarily divisible. By making a decision about the transaction shares for each provider, customers maximize their utility. The sum of all transaction shares gives the entire customer's transaction volume.

In a considered time period, a customer wants to realize **T** transactions in order to satisfy his needs. All **T** transactions represent the entire transaction volume of this time period. Furthermore, the transactions can be carried out by **I** different providers. For such situations, in literature the 'either-or-assumption' is often accepted, in which either all or no transaction can be executed [e.g. 27]. This assumption is, at least in retail markets not very realistic (e.g. in retail banking customers have relations to several financial institutions [4]). For this reason we assume that a customer can choose for each (homogeneous) transaction the best offer (for non-homogeneous transactions see [16]). Thus each provider $i \in I$ carries out transaction shares λ_i.

According to his utility preference function a customer can then determine the directly assigned gross utility value $U_i(\lambda_i)$ for each transaction share and each provider $i \in I$. In addition, for purchase and utilization of the offered services and transactions total costs $K_i(\lambda_i)$ can be calculated. Then, the customers calculates an optimal choice of transaction shares λ_i with $\sum_i \lambda_i = 1$ (executing all T transactions) by maximizing the sum of net utility value $e_i(\lambda_i)$ (gross use value U_i minus costs K_i) over all providers I.

$$e(\lambda) = \sum_i U_i(\lambda_i) - K_i(\lambda_i) \Rightarrow \max!$$
$$\text{with: } \sum_i \lambda_i = 1 \qquad (2)$$

Equation (2) shows a decision situation, in which a customer for example wants to satisfy several, completely *isolated* purchases of fuel. Isolated in this context means that the customers consider only those utility values and costs, which can be directly assigned to a single transaction. With constant utility values and costs for each transaction (i.e. the provider does not change the prices) optimal transaction shares $\lambda_i \in \{0,1\}$ for each provider i result. This corresponds to the above „either-or-premise". But however in reality there are utility values and costs, which are assigned not only to single transaction, but also to several transactions or to the entire business relation. We define such effects as *system effects* and they result from the direct or indirect contact between customer and provider. That means that the customer acts (consciously or unconsciously) in order to benefit for the present or the future by creating utility or by avoiding costs.

System effects can result from different sources. Without going in detail (for details on systems effects see [17]), we can distinguish between systems effects resulted *without intention* and system effects *with intention*. We call system effects created without intended acting by the provider V_H. Even if these system effects are created without intention, they have to be considered in the customer's optimization (2). In contrast, naturally system effects (in further called system effects V_A) are created from providers by *intended acting*. In these situations providers generates goal oriented utility values for customers. As system effects, these utility values are not directly assigned to single transactions, but rather affect a set of subsequent transactions. For example, customer's data can be collected and used for further transactions, in order to enable individualize products [29] or faster execution of identical transactions.

Based on the impact of system effects V_A, we can differ between system effects with a *constant* utility impact and system effects with a continuously *changing* utility impact. System effects V_A with a *constant* utility shall be called $\boldsymbol{V_{A,C}}$. These system effects can be defined depending on a transaction share value in an interval [lower limit **(LL)** $\leq \lambda \leq$ upper limit **(UL)**]. They can span the entire business relation (within the interval of $[0 < \lambda \leq 1]$), as for example created by recommendations for a provider from other customers (reducing the inherent risk). System effects $\boldsymbol{V_{A,C}}$ can also exist, if the transaction share exceeds a certain limit ($\lambda >> 0$). This for example is created by promises of bonus percentage for a number of potential subsequent transactions. As second type, system effects $\boldsymbol{V_{A,V}}$ can have a utility impact, which *changes continuously* depending on the transaction share (change coefficient **v** and exponent γ). An example for this is the possibility to customize services due to gradually collected customer data during previous transaction activity. Again, system effects $\boldsymbol{V_{A,V}}$ could depend on an interval $[LL \leq \lambda \leq UL]$.

If we consider system effects in equation (2) following equation results:

$$e(\lambda) = \sum_i U_i(\lambda_i) - K_i(\lambda_i) + V_{H,i}(\lambda_i) + V_{A,i}(\lambda_i) \Rightarrow \max!$$
$$\text{with: } \sum_i \lambda_i = 1 \tag{3}$$

System effects V_H can be generally defined as

$$V_H(\lambda) = v_H \times (\lambda)^{\mu_H} \text{ with } 0 < \lambda \leq 1 \tag{4}$$

whereby v corresponds to the change coefficient and μ to the exponent of the system effect curve. In contrast to this the system effects V_A can be described as follows:

$$V_A(\lambda) = V_{A,V} + V_{A,C}$$
$$\begin{aligned} \text{with: } & V_{A,V} = v_A \times (\lambda)^{\mu}_{A} && \text{and } LL\ (>0) \leq \lambda \leq UL\ (\leq 1) \\ & V_{A,C} = \text{const.} && \text{and } LL\ (>0) \leq \lambda \leq UL\ (\leq 1) \end{aligned} \tag{5}$$

Having explained and defined system effects, we now introduce a simple example. Later this example is continued to illustrate our results and show the effect of data quality investments in relationships. In our example assume two providers, which are described by following characteristics:

Provider 1:

$U_1(\lambda_1) = 14{,}95\lambda$
$K_1(\lambda_1) = 6{,}8\lambda$

$V_{H(1)}(\lambda_1) = 1{,}2\lambda^{0,5}$
$V_{A,C(1,1)}(\lambda_1) = -1{,}15$ for $0 < \lambda_1 \leq 1$

Provider 2:

$U_2(\lambda_2) = 12{,}5\lambda$
$K_2(\lambda_2) = 4{,}95\lambda$

$V_{H(2)}(\lambda_2) = 1{,}2\lambda^{0,5}$
$V_{A,C(2,1)}(\lambda_2) = -1{,}6$ for $0 < \lambda_2 \leq 1$
$V_{A,C(2,2)}(\lambda_2) = 0{,}8$ for $0{,}8 \leq \lambda_2 \leq 1$

Due to our assumption of homogeneous customer needs and transactions, the utility functions U_1 and U_2 as well as the cost functions K_1 and K_2 have a linear gradient (constant utility values and unit costs for single transactions). In addition, we can assume a concave and for both providers identical gradient for the system effects $V_{H(1)}$ and $V_{H(2)}$. We also have to consider systems effects $V_{A,C(1,1)}$ and $V_{A,C(2,1)}$ as well as costs of preparation the business relation. Provider 2 provides positive system effects $V_{A,C(2,2)}$ by a unique loyalty bonus, which the customer receives only for a transaction share $\lambda_2 \in [0.8;1]$. Other positive or negative systems effects will not be considered.

First we consider directly assignable, isolated net utility value (see equation (1)). We also consider only system effects $V_{H(1)}$ and $V_{H(2)}$, which providers do not influence. All other system effects remain unconsidered. As result we can calculate following transaction shares $\lambda^*_1 = 0.59$ and $\lambda^*_2 = 0.41$ (with $\lambda^*_1 + \lambda^*_2 = 1$):

$$
\begin{aligned}
e(\lambda_1) &= U_1(\lambda_1) - K_1(\lambda_1) + V_{H(1)}(\lambda_1) + U_2(1-\lambda_1) - K_2(1-\lambda_1) + V_{H(2)}(1-\lambda_1) \qquad \Rightarrow \max! \\
&= 14{,}95\lambda_1 - 6{,}8\lambda_1 + 1{,}2\lambda_1^{0{,}5} + 12{,}5(1-\lambda_1) - 4{,}95(1-\lambda_1) + 1{,}2(1-\lambda_1)^{0{,}5} \\
\partial e/\partial\lambda_1 &= 14{,}95 - 6{,}8 + 0{,}6\lambda_1^{-0{,}5} - 12{,}5 + 4{,}95 - 0{,}6(1-\lambda_1)^{-0{,}5} \overset{!}{=} 0 \\
&\rightarrow \lambda_1^* \approx 0{,}59 \ \wedge \ \lambda_2^* \approx 0{,}41 \qquad \text{(with } \partial^2 e/\partial\lambda_1 < 0\text{)}
\end{aligned}
$$

If we explicitly consider the preparation costs and the loyalty bonus of provider 2 (equation (2)), four interval-defined net utility functions *e* exist:

- for $\lambda_1 = 0$ (with $V_{A,C(2,1)}$, and $V_{A,C(2,2)}$)
- for $0 < \lambda_1 \le 0{,}2$ (with $V_{A,C(1,1)}$, $V_{A,C(2,1)}$, and $V_{A,C(2,2)}$)
- for $0{,}2 < \lambda_1 < 1$ (with $V_{A,C(1,1)}$ and $V_{A,C(2,1)}$)
- for $\lambda_1 = 1$ (only with $V_{A,C(1,1)}$).

In this situation the customer dramatically shifts its shares to $\lambda^*_1 = 0{,}2$ and $\lambda^*_2 = 0{,}8$ and thus the transaction share of provider 2 increases to 80%. This situation is illustrated in Figure 1.

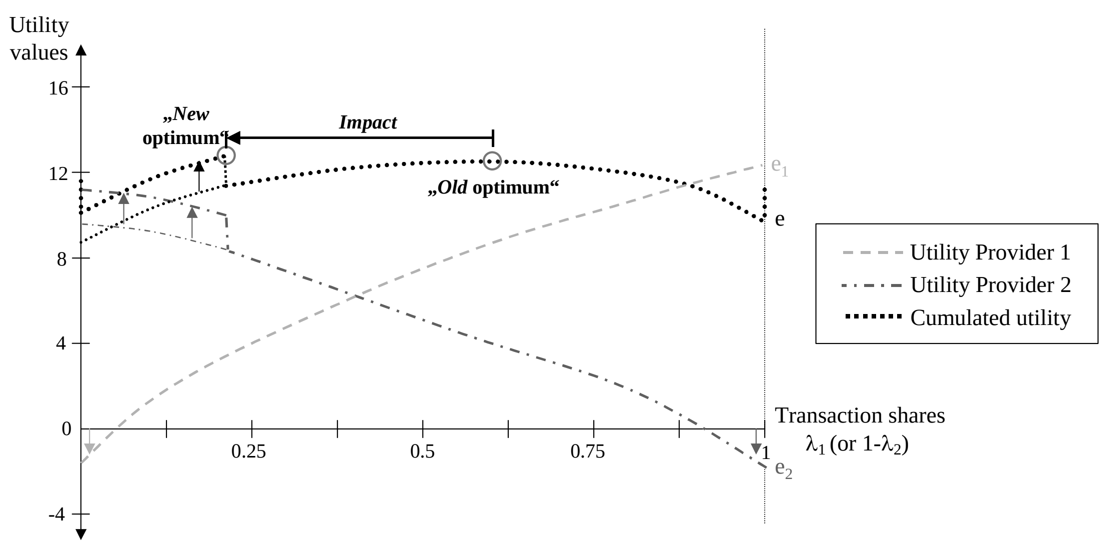

Figure 1: Graphical representation with two-providers

Based on our observation, we now can define customer relationships. In summary, system effects do not aim to improve the utility of a single, isolated transaction in relation to a competitive offer. But these effects "honor" a more intensive and longer lasting business relation, because for example future transactions will be stimulated. For this reason we define relationships as (see also [17]):

A relationship is established as part of the interaction between a customer and a provider (from the customer's view) due to an execution of at least two use-donating transactions or contacts, whereby a subsequent transaction results in particular by the existence and relevance (not necessarily dominance) of provider-generated system effects V_A.

The relevance of provider-generated system effects V_A (sufficient criterion for a relationship) exists especially for the following case: a customer selects an inferior offer (regarding its net utility calculus of isolated transactions and not intended system effects $\boldsymbol{V_H}$) because the utility discrepancy is overcompensated by system effects V_A (character of the relationship). In our example this applies to provider 2 and his transaction shares λ^*_2 increases from 0.41 to 0.8. However, if system effects cannot create any additional transactions (compared to the situation without any system effects), the provider's measures for single transactions are dominant (independent on the height of the system effects), and thus the entire interaction is characterized as transaction-oriented.

5. DATA QUALITY EFFECTS IN CUSTOMER RELATIONSHIPS

Chapter three and four build the foundation for the following section, in which we discuss the question whether data quality influences customer relationships and how it can be represented by a model. First in chapter three we introduced two variables representing improvements of the quality of conformance $Q_t^{conform}$ (see formula (1)): The variable $D_{SUP} \in [0;1]$ represents the data supply and the variable $D_{QM} \in [0;1]$ represents measures as part of reactive and proactive data quality management. In addition, in chapter four we discussed the relevance of system effects as sufficient criteria for customer relationships. Now in this chapter, we address the question how system effects can be created by data quality measures.

As shown, in an initial situation the customer reaches an optimum, were he will not provide any additional data. In this situation, the provider would have to satisfy additional customer's needs to gather any additional data (e.g. offer some additional value). In this situation, the provider could try to increase the quality of conformance by some suitable measures and due to 'better knowledge' about the customer the provider could provide more need-adequate sales recommendations, individual products or convenient execution of business transactions. In this situation the provider has created system effects $V_{A,V(DQ)}$. Initially we consider measures of reactive data quality management (in particular data cleansing), since proactive measures is targeted at subsequent transactions. These measure act as initial investment.

Formally, the system effects created by measures of data quality management are represented by the function

$$V_{A,V(DQ)} = a \times \lambda \times D_{SUP}^{\alpha} \times D_{QM}^{\gamma} \qquad (6)$$

Parameters a, α und γ depend thereby as a function of customer types. These parameters indicate, how a customer (type) perceives better data of his person (e.g. how he appreciates data used for sales recommendations). Following can be noted: As in chapter three discussed the customer's marginal utility decreases monotonic with increasing data supply D_{SUP} ($\rightarrow \alpha \in (0;1)$). This is based on the fact that by completing an existing small database usually a customer perceives new data as relative high value (The provider knows obviously more about the customers). In contrast, the further completion of an already large database, results in lower marginal utility.

But however, the customer still does not provide any further data (due the customer is still in his optimum). D_{SUP} remains (so far) constant, and thus system effects $V_{A,V(DQ)}$ can not be created. For this reason D_{QM} has to be increased initially from the provider by measures of reactive data quality management. A good example is the correction of obviously incorrect customer data by data cleansing measures. These measures are represented by the variable D_{QM}, which has a decreasing marginal utility ($\rightarrow \gamma \in (0;1)$). This is due to the reason that customers perceive data quality improvements in respect of existing data quality. Having initially improved the quality of conformance by reactive data quality management, additional utility in the sense of system effects could (possible!) result and would subsequently result in further transactions. This effect is considered in (6) as transaction shares λ. Due to the premise of homogeneous transactions and its constant utility per transaction, we can omit an additional parameter for the transaction shares λ.

So far in this scenario data quality improvements result only in system effects $V_{A,V(DQ)}$. How thereby the transaction shares increases, has to be examined by means of the relevance of system effects. For studying the relevance of system effects, a minimum level of system effects $V_{A,V(DQ)}$ have to be determined, by which the current transaction shares increases to a new value λ^*. In further, this level is called **level of significance**. If the increasing in system effects is below the level of significance, data quality has no influence on the business relation. In principal, the level can be determined by relating system effects $V_{A,V(DQ)}$ to the optimal transaction shares λ^*. Formally, in the case of a continuous function, equation (3) can be derived in respect to λ. Then, for determine the optimal transaction shares λ^* the derivative $\partial e/d\lambda$ should be zero. For the sake of simplicity we combine constant variables to the parameter *b*, since only $V_{A,V(DQ)}$ is examined and all other variables ($U(\lambda)$, $K(\lambda)$, $V_H(\lambda)$ *and* $V_A(\lambda)$) are constant. Equation (7) represents the general form of this interdependences for the continuity range (e.g. $V_{A,V(DQ)}$ > **level of significance** with one point of discontinuity):

$$\lambda^* = b \times V_{A,V(DQ)}^{\beta} \tag{7}$$

Figure 2 shows a typical illustration for this situation, which is demonstrated with the fact that data quality measures related system effects $V_{A,V(DQ)}$ have to be created at a minimum level of 0.59. As a result the previous transaction shares $\lambda^*_{(old)}$ of 0.2 increases to the new shares $\lambda^*_{(new)}$ of 0.67. If $V_{A,V(DQ)}$ is below 0.59, the transaction shares remains at the old level of $\lambda^*_{(old)}$ = 0.2, although an additional customer's utility is provided. In particular, interval-defined utility functions show interval-depended system effects V_A with two or more (local) maxima and therefore a level of significance exist. In this situation the new utility value at λ^*_{new} differs from the previous global utility maximum (at λ^*_{old}). This is exactly the situation, which providers are able to and want to achieve.

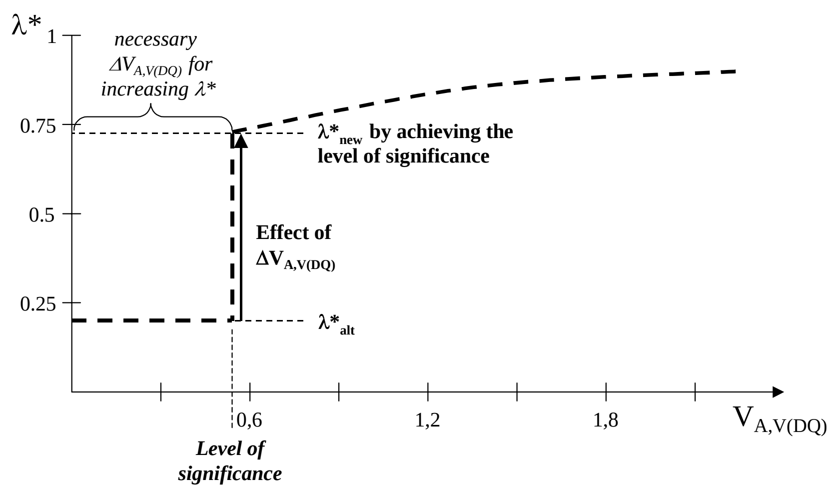

Figure 2: Illustration of the level of significance with data quality measures

The discussion in chapter 3 shows, that data supply D_{SUP} in particular is depended on the transaction shares λ (if we do not consider the option of external data sources). Thus, if the transaction shares λ^*_{old} increases to λ^*_{new}, then additional and/or more current customer data can be acquired (→ positive influence on data quality criteria completeness, timeliness and/or correctness). Representing this formally, the customer data shall be given by the function

$$D_{SUP} = c \times (\lambda^*)^{\sigma} \qquad (8)$$

Parameters c and σ represent the customer and/or transaction's type. Here a concave shape of the function can be justified (→ $\sigma \in (0;1)$). It can be argued, that transactions at a small transaction share contribute more new customer data than transaction already at a high transaction share (especially with homogenous transactions). At this situation we have now an increased transaction share λ^*_{new} and consequently an increased $D_{SUP.}$ This, as equation (6) shows, results in further system effects $V_{A,V(DQ)}$ related to the increased transaction share λ^*_{new}. We call this finding a **feedback effect** or **data quality multiplication effect**. The effect is initiated by increased transaction shares followed by increased customer data, which positively affects the transaction shares again. This feedback continues until the created system effects are below the level of significance.

For illustrating the influence of data quality on customer relations, we continue the example of chapter four. Considering system effects, let us assume the customer determines initially an optimum at $\lambda_1^* = 0.2$ and $\lambda_2^* = 0.8$. Now, provider 1 decides to (initially) invest into data quality, by accomplishing data cleansing measures on existing customer data (e.g. spell checking for address data). Prior we assumed that provider 1 was initially purely transaction oriented and therefore the customer data was not used. Data gathered from previous transactions are not perceived as valuable for relationships and so not used in further transactions (i.e. transaction histories of customers, which are stored in a data base but initially not used for customer contacts). Accordingly measures for neither reactive nor proactive data quality management (e.g. data enrichment) are implemented. Formally, this situation is reflected in $D_{QM} = 0$, in which no system effects $V_{A,V(DQ)}$ are created.

For the example, let us assume that the accomplished investment in (reactive and proactive) data quality measures results in an increase of D_{QM} from 0 to the value of 0.5. In addition, provider 1 decides to use the customer data. Let us also for example set the parameters a to 3, α to 0.8 and γ to 0.65. The situation can then be formally represented as

$$V_{A,V(DQ)} = 3 \times \lambda \times D_{SUP}^{0,8} \times D_{QM}^{0,65}$$

Besides this, representing the creation of customer data as a function of the transaction shares, the equation (8) is of the following form:

$$D_{SUP} = (\lambda^*)^{0,5} \qquad (8)$$

The parameter σ is set to 0.5, which represents a decreasing marginal utility for increasing shares of homogeneous transactions. We also assume that we exclude the option of additional external data sources. Following this assumption, we can conclude the implication of $\lambda^* = 1 \Rightarrow D_{SUP} = 1$ and thus the parameter c in equation (8) must be set to the value 1. Considering the previous customer's optimal transaction shares of $\lambda_1^* = 0.2$ (prior the data quality investment) D_{SUP} results as $0.2^{0,5} = 0.447$. Having invested in data quality and increased thereby D_{QM} to 0.5, provider 1 creates system effects $V_{A,V(DQ)} = 1.004$. Considering the customer optimization in (3) and the created system effect $V_{A,V(DQ)}$ we can calculate for the interval $\lambda_1 \in]0,2;1[$ the a new customer's optimization as:

$$e(\lambda_1) = 14{,}95\lambda_1 - 6{,}8\lambda_1 + 1{,}2\lambda_1^{0,5} - 1{,}15 + \underline{1{,}004\lambda_1} + 12{,}5(1\text{-}\lambda_1) - 4{,}95(1\text{-}\lambda_1) + 1{,}2(1\text{-}\lambda_1)^{0,5} - 1{,}6 \Rightarrow \max!$$

From the customer perspective this equation provides a new optimum at $\lambda_1* = 0.713$ and $\lambda_2* = 0.287$. Due to data quality created system effects, the transaction shares of provider 1 increases ($\Delta\lambda_1* = 0{,}513$). It should be noted, that the used data are provided by the prior transaction shares of λ_1* = 0.2, which also (potentially) effect further transactions. Indeed for this reason the data quality effect is a system effect and as such the level of significance should be consider. To illustrate this, let us assume that provider 1 would invest in data quality and achieves for D_{QM} instead of 0.5 less than 0.22. In this situation, the level of significance for $V_{A,V(DQ)}$ of 0.59λ could not to be exceeded (derived from (7) and see in figure 2) and the customer's total utility would still be at a maximum for transaction shares $\lambda_1* = 0.2$ (or respectively $\lambda_2* = 0{,}8$). Consequently, the data quality measures are without any effect and thus transaction shares for provider 1 would remain the same (despite his investment in data quality measures).

Continuing our example, in the next step the new transaction shares λ_1* of 0.713 leads to an increase of D_{SUP} from 0.447 to $0.713^{0,5}$ = 0.844 without any further intervention of provider 1. The customer perceives this additional created system effects $V_{A,V(DQ)}$, which results in additional customer data. The increased customer data D_{SUP} of 0.844 result in $V_{A,V(DQ)}$ of 1.67, which again feed back to the a new optimal transaction shares λ_1*. Above we described this feedback already as the multiplication effect of data quality. The total effect resulted by the data quality investment are summarized in table 1. It shows that with a single initial investment the transaction shares of λ_1* converges to the value 0.776. The difference to the initial value of λ_1* =0.2 results from the initial increasing of D_{QM} to $\Delta\lambda_1*$ = 0.513 and from the subsequent increasing of D_{SUP} to $\Delta\lambda_1*$ = 0.063. Again, as the example shows, a reciprocal effect between an increased data quality and an improved relationship situation can be explained.

Step	D_{QM}	D_{SUP}	$V_{V(DQ)}$	λ_1*	λ_2*
0	0	0.447	–	0.2	0.8
1	0.5 (Investment)	0.447	1.004λ	0.713	0.287
2	0.5	0.844	1.67λ	0.772	0.228
3	0.5	0.879	1.724λ	0.776	0.224
4	0.5	0.881	1.728λ	0.776	0.224

Table 1: Effects of investments in data quality (example)

6. THE OPTIMAL EFFECTIVENESS OF DATA QUALITY MEASURES

The model illustrated in chapter 5 analyzes the effect of data quality measures in respect to establish and improve customer relationships. So far we didn't address the question, to what extent investments in data quality measures should be taken. In order to normative adjusts the investment decision we address this question and focus on effectiveness of data quality (not efficiency!). Therefore we consider the following further assumption for the <u>provider's</u> calculus:

(P6) A provider determines investments in data quality by effectiveness maximization, i.e. the maximum (quantitative) ratio between the increase of the transaction share $\lambda*$ and the increase of the data quality intensity D_{QM}. Initially in the first step, we do not consider benefits and cost values.

For optimizing the effectiveness, it is necessary to determine the functional dependency between the transaction share $\lambda*$ and the data quality intensity D_{QM}. This is done by the following two steps:

- Combination of equation (7), which represents the functional dependence between the transaction share λ^* and the system effects $V_{A,V(DQ)}$ and equation (6), which represents the system effects $V_{A,V(DQ)}$ as function of the customer data $D_{SUP.}$ ($\Rightarrow \lambda^* = f(D_{SUP}, D_{QM})$).
- The multiplication effect has to be included by a suitable, mathematical model.

Accomplishing the first step we combine (6) and (7) and substitute D_{SUP} with equation (8), which results under consideration of temporal dependence (D_{SUP} results from the optimal transaction share λ_{t-1}^* of the last sub period!) as

$$\lambda_t^* = b \text{ x } (a \text{ x } ((\lambda_{t-1}^*)^{\sigma})^{\alpha} \text{ x } D_{QM}^{\gamma})^{\beta} \Leftrightarrow \lambda_t^* = b \text{ x } a^{\beta} \text{ x } (\lambda_{t-1}^*)^{\sigma\beta\alpha} \text{ x } D_{QM}^{\beta\gamma}$$

For the sake of simplicity, we substitute $m = b \text{ x } a^{\beta}$, $\eta = \beta \text{ x } \gamma$ and $\theta = \sigma \text{ x } \beta \text{ x } \alpha$ and formulate a new equation:

$$\lambda_t^* = m \text{ x } D_{QM}^{\eta} \text{ x } (\lambda_{t-1}^*)^{\theta} \qquad (9)$$

Equation (9) shows a difference function and can be similarly modeled for all other sub periods, i.e. $\lambda^*_{t-1} = f(D_{QM}, \lambda^*_{t-2})$, $\lambda^*_{t-2} = f(D_{QM}, \lambda^*_{t-3})$ etc.; this shall be written as λ^*_{t-1}, λ^*_{t-2} etc. If we combine these functions with (9), the following function (10) results for the sub periods t =:1...T:

$$\lambda_T^* = \prod_{t=1}^{T} (m \text{ x } D_{QM}^{\eta})^{\theta^t} \qquad (10)$$

With $t \to \infty$ function (10) can be represented as geometric series (convergence of the multiplication effect), which can be simplified due to θ (= σ x β x α) < 1 and $t \to \infty$ as:

$$\lambda_T^* := \lim_{t \to \infty} \lambda_t^* = (m \text{ x } D_{QM}^{\eta})^{\frac{\theta^t - 1}{\theta - 1}} \Rightarrow \lambda_T^* = (m \text{ x } D_{QM}^{\eta})^{\frac{1}{1-\theta}} \qquad (11)$$

Equation (11) represents the functional dependency between the transaction share λ^* and the data quality measures D_{QM}. Now we analyze its derivatives $\partial\lambda_t^*/\partial D_{QM}$ and $\partial\lambda_t^*/\partial^2 D_{QM}$ for interpreting the functional characteristics.

$$\frac{\partial\lambda^*}{\partial D_{QM}} = \frac{(m \text{ x } D_{QM}^{\eta})^{\frac{1}{1-\theta}} \text{ x } \eta}{D_{QM} \text{ x } (1 - \theta)} \qquad (12)$$

$$\frac{\partial\lambda^*}{\partial^2 D_{QM}} = \frac{(m \text{ x } D_{QM}^{\eta})^{\frac{1}{1-\theta}} \text{ x } \eta \text{ x } (-1 + \eta + \theta)}{D_{QM}^2 \text{ x } (-1 + \theta)^2} \qquad (13)$$

From $0 < \theta$ (= σ x β x α) < 1, $0 < \eta$ (= β x γ) < 1, $m > 0$ and $0 < D_{QM} < 1$ follows $\partial\lambda_t^*/\partial D_{QM} > 0$ within the continuity range (exceeding the level of significance), i.e. the term (11) is a monotonic increasing function. Thus for an increasing intensity of the data quality measures D_{QM} follow an increase of the optimal transaction share λ^* of $\partial\lambda_t^*/\partial D_{QM}$.

In order to determine the maximum of $\partial\lambda_t^*/\partial D_{QM}$ and thus the highest effectiveness, we analyze the gradient of the function with the second derivative $\partial\lambda_t^*/\partial^2 D_{QM}$:

Due to (12) and (13) follows that (11) is a monotonic increasing function with[1]

1. a concave trajectory for $\eta + \theta < 1$. Because of the decreasing marginal function within the continuity range we can conclude, that the highest effectiveness of data quality measures D_{QM} is located at the point(s) of discontinuity of function (11). Thus in order to determine the point with the highest effectiveness, for $\eta + \theta < 1$ the effectiveness of each point of discontinuity has to be computed and compared.
2. a convex trajectory for $\eta + \theta > 1$. Because of the increasing marginal function within the continuity range, the effectiveness increases continuously with increasing data quality measures D_{QM}. If there exists exactly one point of discontinuity, than the maximum effectiveness results from the maximum intensity of the data quality. If there exist more than one point of discontinuity, the effectiveness of each right-hand limit in the points of discontinuity has to be computed and compared.
3. a linear marginal function for $\eta + \theta = 1$. This means that the effectiveness remains constant with changing data quality intensity.

In summary, in order to achieve a maximum effectiveness providers have to analyze the expression [η + θ </=/> 1]. We illustrate this with an example, but first let us interpret the parameters.

(a) The paramter θ (= σ x β x α) focuses on the customer data D_{SUP}. It includes the data within customer contacts (parameter σ) as well as the creation of system effects (parameter α) and further transaction shares λ^* (parameter β). The parameters σ and α can be directly influenced by the provider. Parameter α can also be influenced, because it considers the use of customer data by providers. Exemplary we refer to customer models, which are developed in [2] and [13] and contain data about the customer, its family and job as well as its attitudes. In contrast, parameter σ describes the gathering and extraction of customer data in transactions, i.e. the creation (not the use!) of the above mentioned customer model. If we would differentiate here between reactive and proactive measures (see assumption (P4)), then σ could be primarily increased by proactive measures. In a long-term parameter σ can be improved by investments in institutionalizing and optimization of such gathering and extraction processes.

(b) In contrast to θ, η (= β x γ) has direct effects on data quality measures (parameter γ) und thus on the creation of system effects. The provider can improve the parameter γ. For this the customer has to realize, that improved data quality results in better services. The customer model, mentioned above, enables to discover (syntactical and semantical) inconsistencies.

(c) As term [η + θ </=/> 1] clearly shows isolated quality improvement -especially if others factors are ignored- have little effects. Consequently a holistic and comprehensive perspective has to be emphasized. This is often stated in data quality management literature [e.g. 11, 18, 34]. In particular this implication is shown by examine the multiplication effect. Increasing $\eta + \theta$ (and thus α, γ and σ) results in higher convexity of function (11) and so an increased multiplication effect, i.e. additional transaction shares.

Finally we illustrate now our findings, by continuing our example.

So far provider 1 specified his data quality measures D_{QM} at a value of 0.5, which resulted in transaction share λ^* of 0.776. Now, due to the functional dependence $\lambda_t^* = f(D_{QM})$, we can calculate the same using equation (11). In chapter 5 we set the parameters b and β of equation (7) as b = 0.715843 and β = 0.146158, and thus we realize for our example following results:

$$m = b \times a^{\beta} = 0.715843 \times 3^{0.146158} = 0.84053 \qquad \eta = \beta \times \gamma = 0.146158 \times 0.65 = 0.095$$

[1] That the specifications 2. and 3. (increasing or linear marginal functions) have to be considered as special cases. In our article it is formulated as thesis and shall be discussed in detail in future research.

$\theta = \sigma \times \beta \times \alpha = 0.5 \times 0.146158 \times 0.8 = 0.05846$ $D_{QM} = 0.5$

Using these values in equation (11) we calculate $\lambda^* = 0.775$. This corresponds approximately (due to rounding errors) to the converged transaction share in table 1. But is $D_{QM} = 0.5$ effective in the sense of assumption (P5)? For providing an answer, we have to analyze the equation (11), i.e., whether $\eta + \theta$ is [</=/>] 1. In the example $\eta + \theta < 1$ for $\eta = 0.095$ und $\theta = 0.05846$ results, from which a concave trajectory and a decreasing marginal function of $\lambda^* = f(D_{QM})$ can be concluded. For this reason the maximum efficiency exists on a point of discontinuity of function (11). Because provider 1 do not offer any interval-defined system effects for $0 < \lambda^* < 1$ (e.g. no loyalty bonus), only the level of significance is the point of discontinuity. The level of significance was determined with $D_{QM} = 0.22$ in chapter 5. If we consider $D_{QM} = 0.22$ in function (11), we receive as result $\lambda^* = 0.711$.

The example shows that the reduction of D_{QM} from 0.5 to 0.22 (about 56%) results in a decline of the transaction share from $\lambda^* = 0.775$ to $\lambda^* = 0.711$ (only about 8%). Finally, table 2 shows the development of the effectiveness $\Delta\lambda^*/\Delta D_{QM}$ and the changing transaction share $\Delta\lambda^*$ (Starting point is $\lambda^* = 0.2$) for alternative definitions of D_{QM}:

ΔD_{QM}	<0.22	0.22	0.35	0.5
$\Delta\lambda^*$	0	0.511	0.547	0.575
$\Delta\lambda^*/\Delta D_{QM}$	0	2.32	1.56	1.15

Table 2: Effectiveness $\Delta\lambda^*/\Delta D_{QM}$ for alternative data quality measures (example)

Without regarding to the difficult practical measurement of the parameters as well as the problem of the (different) scales of λ^* and D_{QM}, the example shows nevertheless impressively the fact that the definition of data quality measures is substantially more multilayered than statements like “data quality is per se useful in the CRM”. The next paragraph summarizes the results of our research briefly and points out some implications for practice and further research.

7. CONCLUSIONS AND FUTURE RESEARCH

In literature many authors pose the importance of data quality for CRM and therefore assuming a positive correlation between data quality and relationships [e.g. 23; 7]. Our research contributes to this research area and analyses interdependencies between data quality investments and customer relationships. In this article we intended to explain effects of data quality investments to customer relationships. Conceptualizing data quality and the construct of customer relationship, we developed a model. Based on this model, we analyzed the question, whether and in which cases investments in data quality intensify relationships. Our findings show, in contrast to transaction-oriented interactions, how data quality can be used to create systems effects. However, as our results also stress, data quality do not necessarily lead to improved business relationships.

First of all, it requires a customer’s affinity for data quality, i.e. the customer perceives utility of storing and using his data. In addition, in order to intensify relationships the created utility has to exceed a level of significance. Both aspects are initiated by CRM and can be used as important control factors for data quality management in CRM. Based on our results, conclusions for practice can be derived, like for example cost effectiveness considerations for data quality management. Indeed, the level of significance must be exceeded, but at the same time aiming for an extremely high data quality is under economical considerations questionable. Besides this, the results show the so-called multiplication effect of data quality, which is created by the feedback of transaction shares and the generated data supply. In contrast to other relationship values, like confidence or monetary incentives [see 16], system effects are multiplied by the qualitatively higher customer data. In order to analyze the impact of proactive data quality measures for creating such feedbacks, the model provides first results and should be studied further. But however, finally the developed model has to be critical reviewed and further discussed. Following we summaries some critical points, which should be addressed in future research:

1. Indeed, the underlying model premises are a critical aspect of the model. On the one hand this was necessary to establish a suitable theoretical basis. On the other hand the premises limited the model and its context should be extended in future research. In particular it should be extended to heterogeneous transactions (e.g. different bank products) as well as incorporate dynamic aspects (consider that providers make mutually dependent decisions).

2. Further research should concretizing and empirical validate the functional dependencies between data quality and customer relationships (e.g. parameters for different customer types). Further research should also validate and estimate the defined variables, in particular data supply D_{SUP}, data quality measures D_{QM} and system effects $V_{A,V(DQ)}$. In order to estimate the level of significance and the multiplication effect, further studies on system effects are necessary. Realistically it can be assumed, that different customer types have different levels of significance. This would raise the question, with which intensity data quality measures have to be taken? But since data quality measures in practice can usually not be selectively implemented for individual customers, an adequate intensity for these measures has to be defined. On the one hand, this is necessary in order to ensure that many customers exceed the level of significance and thus creating an effect. But on the other hand, with intensive measures the cost of resources might not be covered by the associated benefits. Therefore, a number of interesting, practice-relevant considerations arise, which should be addressed by further empirical research (e.g. conjoint analysis for the estimation of data quality benefits in terms of system effects).

3. Efficiency of data quality measures is not addressed in our research so far. Further research should address this, in addition to aspects of effectiveness. Research should study the cost/benefit ratio between customer data as non-monetary value and other relationship values (e.g. provider's bonus promises).

REFERENCES

[1] Arndt, D., Langbein, N. Data Quality in the Context of Customer Segmentation. In, Fischer, C., Davidson B. (eds.) *Proceedings of the 7th International Conference on Information Quality*, Cambridge 2002, pp. 47-60.

[2] Buhl, H. U., Fridgen, M., Volkert, S. Systemunterstützte individualisierte Kundenansprache in der Mehrkanalwelt der Finanzdienstleistungsbranche – Repräsentation der Einstellungen von Kunden in einem Kundenmodell. In, *Proceedings of the 6th International Conference for Business Informatics 2003*, Dresden 2003.

[3] Cappiello, C., Francalanci, Ch., Pernici, B., Plebani, P., Scannapieco, M. Data Quality Assurance in Cooperative Information Systems: A multi-dimensional Quality Certificate. In, Catarci, T. (ed.): *International Workshop on Data Quality in Cooperative Information Systems*, Siena 2003, pp. 64-70.

[4] Daniel, J. *Ertragssteigerung durch Kundenzufriedenheit und Kundenbindung. Möglichkeit und Grenzen am Beispiel eines Kreditinstituts*. Dt. Sparkassen-Verlag: Stuttgart, 2001.

[5] Diller, H. Kundenbindung als Marketingziel. *Marketing ZFP*, 18 (2), 1996, pp. 81-94.

[6] Deutsches Institut für Normung e. V. *Qualitätsmanagement, Statistik, Zertifizierung: Begriffe aus DIN-Normen*. 2nd ed., Beuth: Berlin et al., 1995.

[7] Dubois, L. New Channels Present New CRM Challenge. *The American Banker*, 7. Feb. 2001.

[8] Eager, A. CRM data needs to come clean. *Computer Business Review*, 9 (11), 2001, pp. 40.

[9] Eckerson, W. *Data Quality and the Bottom Line*. The Data Warehousing Institute Report Series No. 101, Chatsworth 2002.

[10] Eggert, A. *Kundenbindung aus Kundensicht: Konzeptualisierung – Operationalisierung – Verhaltenswirksamkeit*. Gabler: Wiesbaden, 1999.

[11] English, L. *Improving Data Warehouse and Business Information Quality*. Wiley: New York, 1999.

[12] Eriksson, K., Fjeldstad, O. Value Creation in Financial Services, "The future of retail banking and customer relationships". In, *The first workshop of The Göran Collert Research Foundation*, Sep. 13-14, 2001.

[13] Fridgen, M, Steck, W. *New Perspectives on Individualization and Controlling of Web Sites for the Financial Services Industry*. working paper WI-99, Department of Information Systems, University of Augsburg, 2001.

[14] Garvin, D. A. What does "Product Quality" really mean? *Sloan Management Review*, 26 (1), 1984, pp. 25-43.

[15] Grönroos, C. The relationship marketing process: interaction, communication, dialogue, value. In, *2nd conference on relationship marketing*, Monash University, 1999.

[16] Heinrich, B. *Methode zur wertorientierten Analyse und Gestaltung der Kundeninteraktion – Zur Rolle des Service Integrators im Privatkundengeschäft von Kreditinstituten*. Logos: Berlin 2002.

[17] Heinrich, B., Helfert, M. Nützt Datenqualität wirklich im CRM? – Wirkungszusammenhänge und Implikationen. In, *Proceedings of the 6th International Conference for Business Informatics 2003*, Dresden 2003.

[18] Helfert, M. *Proaktives Datenqualitätsmanagement in Data-Warehouse-Systemen. Qualitätsplanung und Qualitätslenkung*. Logos: Berlin 2002.

[19] Helfert, M., Herrmann, C. Proactive data quality management for data warehouse systems. In, *4th Intern. Workshop on 'Design and Management of Data Warehouses' (DMDW'2002) in conjunction with CAiSE 2002*, Toronto, 27. Mai 2002, pp. 97-106.

[20] Hinrichs, H. *Datenqualitätsmanagement in Data Warehouse-Systemen*. doctoral thesis, University Oldenburg, 2002.

[21] Huang, K., Lee, Y., Wang, R. Y. *Quality Information and Knowledge*. Prentice Hall: Upper Saddle River, 1999.

[22] Jarke, M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P. *Fundamentals of data warehouses*. Springer: Berlin et al., 2000.

[23] Khalil, O. E. M., Harcar, T. D. Relationship Marketing and Data Quality Management. *SAM Advanced Management Journal*, 64 (2), 1999, pp. 26-33.

[24] Naumann, F. *Quality-Driven Query Answering for integrated Information Systems*. Springer: Berlin et al., 2002.

[25] Nelson, S. *What's happening to CRM in 2002*. Gartner Group, January 2002.

[26] O'Malley, L., Tynan, C. Relationship marketing in consumer markets – Rhetoric or reality? *European Journal of Marketing*, 34(7), 2000, pp. 797-815.

[27] Plinke, W. Grundlagen des Geschäftsbeziehungsmanagements. In, Kleinaltenkamp, M., Plinke, W. (eds.) *Geschäftsbeziehungsmanagement*. Springer: Berlin et al., 1997, pp. 1-61.

[28] Redman, Th. C. *Data Quality for the Information Age*. Artech House: Boston and London, 1996.

[29] Schackmann, J., Link, H. Mass Customization of Digital Products in Electronic Commerce. In, Sebaaly, M. F. (ed.) *Proceedings of the International NAISO Congress on ISI 2001*, pp.144-150.

[30] Seghezzi, H. D. *Integriertes Qualitätsmanagement: das St. Galler Konzept*. Hanser: Munich, 1996.

[31] Storbacka, K., Strandvik, T., Grönroos, C. Gewinn durch Relationship Management. In, Payne, A., Rapp, R. (eds.) *Handbuch Relationship Marketing: Konzeption und erfolgreiche Umsetzung*. Vahlen: Munich, 1999, pp. 69-89.

[32] Teboul, J. *Managing Quality Dynamics*. Prentice Hall: New York, 1991.

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

[34] Wolf, P. *Konzept eines TQM-basierten Regelkreismodells für ein „Information Quality Management" (IQM)*. Verlag Praxiswissen: Dortmund, 1999.