University of Augsburg
Prof. Dr. Hans Ulrich Buhl
Research Center
Finance & Information Management
Department of Information Systems
Engineering & Financial Management

Universität
Augsburg
University

Discussion Paper WI-341

An Economics-Driven Decision Model for Data Quality Improvement – A Contribution to Data Currency

by

Quirin Görz

in: Proceedings of the 17th Americas Conference on Information Systems, AMCIS.
Paper 145

Universität Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)
www.fim-online.eu

An Economics-Driven Decision Model for Data Quality Improvement – A Contribution to Data Currency

Quirin Görz
FIM Research Center,
Professorship for Business
and Information Systems Engineering &
Management Support
University of Augsburg[1]
Universitaetsstr. 12, Augsburg, Germany
quirin.goerz@wiwi.uni-augsburg.de

ABSTRACT

As poor data quality usually leads to high costs, managing data quality is essential for organizations. Thereby, comparing the current with the required data quality level is necessary for an effective and economics-driven data quality management. Otherwise decision makers might decide in favor of unsuitable or inefficient data quality improvement measures with respect to cost and benefit. Existing methodologies for assessing and improving data quality often neglect providing methods for determining the required data quality level or argue on a managerial rather than an operational level. As a consequence, an economics-driven and context-dependent decision model for updating data at the level of attribute values is presented. This model contains a metric for currency, errors and error costs, and a currency threshold for attribute values. The decision model is illustrated using a direct marketing example.

Keywords

Data quality, data quality management, currency, decision model.

INTRODUCTION

Poor data quality (DQ) usually leads to wrong decisions and high costs (Fisher, Chengalur-Smith and Ballou, 2003, p. 170; Ballou and Tayi, 1999, p. 73). For instance, 75% of the interviewees in an international study on DQ admitted wrong decisions due to incorrect data (Harris Interactive, 2006). Another study conducted by the Data Warehouse Institute revealed that in 67% of the involved organizations poor DQ results in high costs (Russom, 2006, p. 11). Consequently, ensuring DQ is still an important issue for organizations (Russom, 2006; Ballou, Wang, Pazer and Tayi, 1998, p. 462; Jiang, Sarkar, De and Dey, 2007, p. 1946).

Ensuring DQ requires an efficient data quality management (DQM) with respect to cost and benefit. DQM thereby can refer to a managerial or an operational level. At a managerial level, DQM deals with DQ projects, DQ service levels or DQ investments in general (Askira Gelman, 2010; Ballou and Tayi, 1999; Cappiello and Comuzzi, 2009; Heinrich and Klier, 2006). At an operational level, DQM is context-dependent and deals with reconciling attribute values from multiple data sources or conducting particular DQ improvements (e. g. data cleansing) on the level of attribute values (Jiang et al., 2007). Although several metrics for DQ as well as methodologies for assessing and improving DQ on a managerial level exist, methodologies for operational DQM (especially for specific DQ dimensions such as currency) are still missing.

Hence, the following research question arises: *How can a decision maker decide on conducting context-dependent DQ improvements on the level of attribute values?*

[1] Financial support from the Research Center Global Business Management at the University of Augsburg is gratefully acknowledged.

As a first step to answer this research question, a decision model for updating data (as a specific DQ improvement) on the level of attribute values will be developed. This is reasonable because up-to-date data are essential for important business operations such as customer relationship management or supply chain management. Beyond, currency is one of the most cited DQ dimensions (Wand and Wang, 1996, p. 92; Lee, Strong, Kahn and Wang, 2002, p. 134) and is often used synonymously with timeliness (cf. Table 1). Dependencies between different DQ dimensions as discussed in Helfert, Foley, Ge and Cappiello (2009) are subject to further research and thus will be neglected in the following.

Currency can be defined as the characteristic of an attribute value that it is still up-to-date at its valuation date (instant of measurement). That is that an accurately recorded attribute value conforms to currency if it corresponds to its real-world representation at a (later) valuation date and has not become outdated (due to a temporally decline) in the meantime. With respect to the later decision model currency will be defined according to Heinrich, Kaiser and Klier (2007b) as the (conditional) probability that an attribute value is still up-to-date.

Authors	*Definition*
(Ballou et al., 1998; Ballou and Pazer, 1995)	**"The record value is not out of date."**
(Wang and Strong, 1996)	**"The extent to which the age of the data is appropriate for the task at hand."**
(Redman, 1996)	**"Timeliness refers to a degree to which a datum in question is up-to-date. A datum value is up-to-date if it is correct in spite of possible discrepancies caused by time-related changes to the correct value."**
(Pipino, Lee and Wang, 2002)	**"the extent to which the data is sufficiently up-to-date for the task at hand."**
(Batini and Scannapieco, 2006)	**"Timeliness expresses how current data are for the task at hand."**
(Heinrich et al., 2007b; Heinrich, Kaiser and Klier, 2007a)	**" Timeliness can be interpreted as the probability that an attribute value is still up-to-date"**

Table 1 Definitions for currency (selection)

In order to decide on updating an attribute value before its usage in a particular context, the attribute value's current currency, i. e. the value of its metric for currency, has to be compared to a required level of currency, i. e. a currency threshold. An value of an attribute should hence be updated, if its current currency is less than the context-dependent threshold.

The article is structured as follows: At first related work on methodologies for assessing and improving DQ as well as metrics for measuring currency are discussed. Then the decision model is developed and illustrated using a direct marketing example. Results and limitations are summarized at the end.

RELATED WORK

Today several methodologies for DQM exist. An economics-driven DQM usually consists of defining, measuring, analyzing and improving DQ (Wang, 1998, p. 60ff.). In this context it is essential to decide on conducting DQ improvements based on the current as well as the required DQ level (Ballou and Tayi, 1999; Batini, Cappiello, Francalanci and Maurino, 2009; Cappiello and Comuzzi, 2009; Cappiello, Even and Kaiser, 2009; Francalanci and Pernici, 2004; Otto, Ebner and Hüner, 2010; Pierce, 2004; Shankaranarayanan, Ziad and Wang, 2003; Umar, Karabatis, Ness, Horowitz and Elmagardmid, 1999). This implies that the current DQ level should be compared with the required DQ level. An overview on thirteen methodologies for data quality assessment and improvement can be found in Batini, Cappiello, Francalanci and Maurino (2009). Batini et al. (2009) reveal that none of the investigated methodologies provides decision models for conducting specific DQ improvements on the level of attribute values or methods to determine the required DQ level for particular DQ dimensions. Only the assess and improve quality (AIMQ) methodology by Lee et al. (2002) focuses on benchmarking methods to decide on improving the DQ of a database or organization.

Apart from that, several other methodologies exist, dealing with DQ assessment, i. e. comparing the current with a required DQ level, to enable a diagnosis of quality (Ballou and Tayi, 1999; Cappiello and Comuzzi, 2009; Cappiello et al., 2004; Even and Kaiser, 2009; Otto et al., 2010; Pierce, 2004; Shankaranarayanan et al., 2003; Umar et al., 1999).

Ballou and Tayi (1999) as well as Cappiello and Comuzzi (2009) discuss utility-based methodologies to decide on DQ projects or the optimal DQ level in IT service offerings. In Ballou and Tayi (1999) the decision maker has to assess for each project, data set, and relevant DQ dimension the change in utility composed by the current quality, required quality, anticipated quality, priority of organizational activity, cost of DQ enhancement, and value added. In Cappiello and

Comuzzi (2009) the optimal IT service offering – and hence the service's optimal DQ level – is identified as the one that maximizes the sum of the service provider and customers' utility functions. Information product (IP)[2]-oriented methodologies for managing DQ can be found in Shankaranarayanan et al. (2003), Pierce (2004), and Even and Kaiser (2009). Shankaranarayanan et al. (2003) present a framework to manage DQ using a modeling technique to explicitly represent the manufacture of an IP, quality dimensions, methods to compute DQ, and a set of capabilities. Pierce (2004) develops a model based on control matrices, to assess the quality of an IP with respect to several DQ dimensions. Both models allow decision makers not only to understand and evaluate an IP used in a decision task, but also to understand and evaluate its quality. Even and Kaiser (2009) present a framework for the economics-driven assessment of DQ decisions. The objective is to detect the set of DQ improvements which maximizes the net benefit from improvement to the utility gained by the use of IPs, cost added or saved in the information process, and costs attributed to the different stages of the decision process. A so-called "pragmatic approach" for enterprise DQ based on a case study in telecommunication industry can be found in Umar et al. (1999). Thereby a methodology to synthesize DQ issues into an architectural vision is developed, supported by a series of steps, procedures, checklists, and tools. Another case study based approach can be found in Otto et al. (2010). Here a DQ index is compared to a target DQ index in order to decide on further DQ improvements. The current DQ index is therefore computed based on business rules. A model relating the assessment phase to user requirements is suggested by Cappiello et al. (2004). Within this model users can establish individual, subjective limits of acceptability for each quality dimension and service. At the end of the assessment procedure, the user is provided with both the requested data as well as the quality evaluation's results – i. e. the requested data does or does not meet the user's requirements.

All these methodologies have in common that making a decision on conducting DQ improvements is context-dependent, has to be performed with respect to different DQ dimensions, and has to consider user requirements. Furthermore, the current DQ level should be compared with or related to a required DQ level. Nonetheless, the introduced methodologies barely present methods to determine the required DQ level. It is argued that the required DQ level can be determined by the organization's commitment to DQ (Otto et al., 2010; Pierce, 2004) or subjective user requirements/preferences (Cappiello and Comuzzi, 2009; Cappiello et al., 2004). Moreover, only the methodologies presented by Shankaranarayanan et al. (2003), Pierce (2004), Even and Kaiser (2009), Cappiello et al. (2004), and Cappiello and Comuzzi (2009) provide methodologies on an operational level as they refer to IPs or IT services which can also represent single attribute values.

Beside these methodologies several metrics for quantifying the current DQ level exist. Formal metrics for quantifying currency can be found in Ballou et al. (1998), Even and Shankaranarayanan (2007), Heinrich and Klier (2009), and Hinrichs (2002). These metrics – except of the one defined by Heinrich and Klier (2009) – have already been discussed in Heinrich, Kaiser and Klier (2009). Thus, solely the metric presented by Heinrich and Klier (2009) is to be discussed briefly.

Heinrich and Klier (2009) develop a metric for currency based on the following considerations: (1) The metric is based on probabilistic theory and its value can be interpreted as the conditional probability that the considered attribute value ω is still up-to-date. (2) Besides the attribute metadata "date of occurrence" t_0 and "shelf life" T, the metric takes into account so-called supplemental data w_i. Whereas "supplemental data are additional data attributes that allow drawing conclusions about the currency of the data attribute considered" (Heinrich and Klier, 2009). This is why the metric represents the conditional probability that the considered attribute value ω is still up-to-date. Because of its probabilistic interpretation, this metric will be used for the later decision model.

Related work reveals that several formal metrics for quantifying the current DQ level, especially for the DQ dimension currency exist. However, none of the investigated methodologies for assessing and improving DQ contains an intersubjectively verifiable and context-dependent determination of the required DQ level. Beyond, they barely present approaches suitable for particular attribute values. As a first step in order to address this research gap with respect to the DQ dimension currency, an economics-driven decision model for updating data on the level of attribute values will be developed.

DECISION MODEL

The main idea of the decision model is to compare an attribute value's current with its required currency in order to decide on updating the value of an attribute with respect to a particular context. The current currency will therefore be quantified by applying a metric for currency while the required currency will be represented by a threshold. Based on these two values for currency, an attribute value ω should be updated if the current currency, i. e. the probability that the attribute value is still up-to-date, is less than the context-dependent threshold.

---

[2] An information product can be defined as an information output of an information manufacturing system, which has value and is transferred to an internal or external "customer" (Ballou et al., 1998, p. 463).

The decision model consists of three parts described in detail in the following subsections. The first one is a metric for currency based on Heinrich and Klier (2009) in order to determine the current currency. The second one is the definition of errors and error costs for updating or not updating an attribute value. The third one is the development of a method for calculating context-dependent currency thresholds.

Measuring Currency

As a first step towards the decision model the question has to be answered on how to measure the current currency of an attribute value ω. Therefore the metric proposed by Heinrich and Klier (2009) is used. This is reasonable because for determining the currency threshold it is necessary to calculate the expected total costs for not updating an attribute value (see below). Hence, currency is defined as a probability.

The metric defined by Heinrich and Klier (2009) is represented by

$$Q_{curr}^{\omega}(t, w_1,..., w_n) := P^{\omega}(T \geq t \mid W_1 = w_1,..., W_n = w_n) \quad (1)$$

Currency $Q_{curr}^{\omega}(t, w_1,..., w_n)$ is thereby defined as the conditional probability that shelf life T, i. e. the maximum validity of an attribute value, is larger than or equal to the age t of the attribute value ω, given the supplemental data w_i. The supplemental data w_i are a realization of the variable W_i and allow drawing conclusions about the currency of the attribute considered.

This probability represents the current currency for a single attribute value ω which should be compared to its threshold.

So as to calculate this threshold the question appears which kind of errors can occur in association with a decision on updating or not updating an attribute value and what kind of costs arise from these errors. These errors and error costs will be examined next.

Errors and Error Costs

In the context of deciding on updating or not updating an attribute value, two types of errors can be differentiated:

A *type I error* occurs, when an attribute value ω that should have been updated has not been updated. That is an out-of-date attribute value will be used and consequently leads to a wrong decision (e. g. a customer cannot be reached successfully). As a result, costs occur in terms of lost profit. The profit arises from the expected payoff realized by an up-to-date attribute value's usage represented by $X \in IR_0^+$ and costs for an attribute value's usage represented by $Y \in IR_0^+$. Accordingly, the lost profit and hence costs of a type I error equal $(X - Y)$.

A *type II error* occurs, when an attribute value ω which should *not* have been updated is updated. That is an attribute value will be updated unnecessarily due to the fact that it already was up-to-date. As a consequence avoidable costs for updating the attribute value occur. These updating costs are represented by $C \in IR_0^+$. Accordingly, costs of a type II error equal C.

Based on these errors and error costs the question on how to calculate the total costs for updating or not updating an attribute value has to be answered as well as on how to calculate a currency threshold based on these total costs. Both questions will be investigated in the following subsection.

A Currency Threshold for Decision Making

The outcome of the dichotomous decision at hand can be "Update" or "No update". The total (expected) cost for the two decision outcomes can be determined on the basis of the metric for currency and the error cost parameters stated above. The total costs of the two decision outcomes are depicted in Table 2.

Decision	*Type I error costs*	*Type II error costs*
Update	N/A	C
No update	$(X-Y)(1-Q_{curr}^{\omega}(\cdot))$	N/A

Table 2 Costs associated with updating and not updating an attribute value

If the decision is "Update", only a type II error can occur. Since an update will definitely be conducted, the total costs of updating an attribute value are deterministic. The total costs for the decision outcome "Update" thus equal

$$TC^{\omega}(update) = C \quad . \tag{2}$$

If the decision is “No update”, only a type I error can occur. In this case the attribute value is only out-of-date with a certain probability. The resulting costs can hence be interpreted as expected costs. In order to calculate these expected costs, the type I error costs $(X - Y)$ are multiplied with the probability that the attribute value is out-of-date at its instant of usage. This probability can be calculated based on the complementary probability $(1 - Q^{\omega}_{curr}(t, w_1,..., w_n))$ of the metric for currency. The expected total costs for the decision outcome “No update” thus equal

$$TC^{\omega}(no\ update) = (X - Y)\left(1 - Q^{\omega}_{curr}(t, w_1,..., w_n)\right) \ . \tag{3}$$

From an economic point of view, updating an attribute value is only reasonable if the total costs for updating are less than the expected total costs for not updating an attribute value, i. e. $TC^{\omega}(update) < TC^{\omega}(no\ update)$. By applying (2) and (3) the following currency threshold can be derived:

$$Q^{\omega}_{curr}(t, w_1,..., w_n) < \frac{(X - Y) - C}{(X - Y)} \ . \tag{4}$$

If the value of the metric for currency is less than the right-hand side of formula (4), the attribute value ω should be updated.

By doing so the currency threshold stated in formula (4) is only defined for $(X - Y) > 0$ and $((X - Y) - C) \geq 0$. Otherwise the right-hand side of formula (4) – in case $(X - Y) = 0$, division by zero or for $(X - Y) < 0$ or $((X - Y) - C) < 0$, negative probability – is not defined. This holds because for $(X - Y) \leq 0$ and $((X - Y) - C) < 0$ an attribute value’s usage would always generate zero profit or a loss, even if the attribute value is up-to-date. As a consequence, the attribute value’s usage would not be economically reasonable at all.

These two cost parameters for a type I $((X - Y))$ and a type II error (C) have a reverse effect on the threshold. As can be seen from inequality (5), an increase in lost profit from an attribute value’s usage $(X - Y)$ c. p. leads to an increase of the threshold. Moreover, the marginal threshold is decreasing with an increasing lost profit $(X - Y)$ (see inequality (6)). Thus, the higher (less) the costs for a type I error are, the higher (less) c. p. is the resulting threshold and the less (higher) is the marginal threshold.

$$\frac{\partial \frac{(X-Y)-C}{(X-Y)}}{\partial (X-Y)} = \frac{C}{(X-Y)^2} > 0 \tag{5}$$

$$\frac{\partial \frac{C}{(X-Y)^2}}{\partial (X-Y)} = -\frac{2C}{(X-Y)^3} < 0 \tag{6}$$

When examining the updating costs C it can be seen from inequality (7), that an increase in updating costs C c. p. leads to a linear decrease of the threshold. Thus, the higher (less) the costs for a type II error are, the less (higher) is the resulting threshold.

$$\frac{\partial \frac{(X-Y)-C}{(X-Y)}}{\partial C} = -\frac{1}{(X-Y)} < 0 \tag{7}$$

The threshold is hence the higher the more expensive it is to update a particular attribute value in comparison to the potential profit lost due to its usage without an update. These reverse effects are also shown in Figure 1.

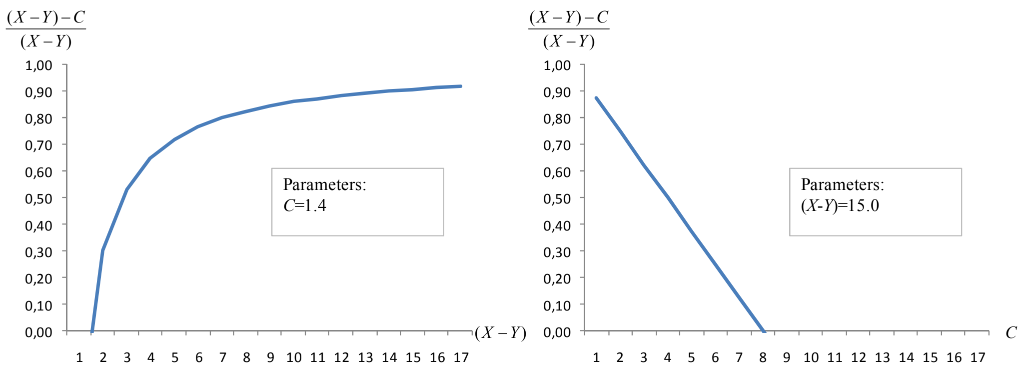

Figure 1. Visualization of the reverse effects of the type I and II error costs on the threshold

To summarize, a decision maker has to take the following four steps in order to decide on updating or not updating an attribute value: (i) calculate the currency metric for each attribute value considered, (ii) determine the cost parameters for a type I and II error, (iii) calculate the currency threshold, and (iv) compare the threshold with each metric result of the attribute value.

A DIRECT MARKETING EXAMPLE

The practical applicability of the decision model stated above will now be demonstrated using a short direct marketing example.

In this case a company is considered which has decided to launch a direct marketing campaign for any product. In order to come into contact with a customer, the company uses customer data, i. e. the customer's address data which is stored in a customer data base. Before launching the campaign the company has the option to update address data by buying a customer's address form postal service operators (e. g. for German customers from Deutsche Post AG). As this would usually be very expensive for all customers, the company could apply the decision model stated above.

As a first step, the company has to quantify the current currency of each customer's address. Therefore, shelf life T of an address is assumed to be exponentially distributed (Heinrich et al., 2007b, p. 8). The metric for currency of an attribute value ω (with ω representing a customer's address) can hence be defined as follows:

$$Q^{\omega}_{curr}(t, w_1,..., w_n) := \exp(-decline(A, w_1,..., w_n)\cdot t)$$

The parameter $decline(A, w_1, ..., w_n)$ represents the decline rate indicating how many values of the attribute A become out-of-date on average within one period of time, considering supplemental data w_i. These supplemental data w_i (e. g. a country, state, or city) influence the decline rate (e. g. for some countries the decline rate might be higher than for others). Based on this metric the company can quantify the currency of an address with age t. Age t can be calculated as the difference between the instant when the attribute value is used (t_1) and the instant of data acquisition (t_0). The instant of data acquisition is sufficient because the exponential distribution is memoryless. For instance, the value of the metric would be 0.90 for a address with $decline(A, w_1, ..., w_n) = 0.05$ and $t = 2$. This quantification can be performed in a cost efficient manner by means of SQL DML-statements.

Afterwards, the company has to determine the cost parameters for a type I and II error. The campaigns expected payoff could for example be assumed to be \$ 10.00 for each successfully reached customer (cost parameter X). Calculated by multiplying the payoff generated by the offer (e. g. \$ 100) and the respective ex ante success rate (e. g. 10%) representing the offer's probability of acceptance. Costs to come into contact with the customer can be assumed to be \$ 0.53 to mail and another \$ 1.50 for the product information materials (cost parameter Y). The cost parameter for a type I error (X - Y) consequently is (\$ 10.00 – (\$ 0.53 + \$ 1.50)) = \$ 7.97. As already stated above the company can buy up-to-date address data from postal service operators. Thus, it is assumed that one address can be bought for \$ 1.40 (cost parameter C). So costs for a type II error C are \$ 1.40.

The currency threshold stated in formula (4) can now be calculated for each address, resulting in a threshold being 82.50%. Thus, the company should buy an address if its currency is less than 82.50%. For instance, the address stated above with a value of the metric being 90% does not have to be updated before its usage within this campaign. This comparison can also be performed by means of SQL DML-statements.

Since the considered cost parameters vary for different campaigns and currency also decreases over time, a decision maker has to calculate the threshold and the value of the metric for each campaign anew. However, calculating currency can be performed repetitively in an automated manner. As a consequence, solely estimating the cost parameters generates recurrent costs. As these costs have to be estimated anyway in the course of campaign planning, costs arising from applying the decision model can be neglected after it has been implemented once.

As can be seen from this example the decision model can be applied to direct marketing campaigns. Nonetheless, this is just an example.

SUMMARY

In this paper an economics-driven decision model for updating data on the level of attribute values has been developed. The model's objective is a context-dependent identification of attribute values, which should be updated before their usage based on currency thresholds. The presented model contains the following parts: A DQ metric for the DQ dimension currency, errors and error costs, and a currency threshold. All together they build a decision model which enables a decision maker to decide on updating or not updating an attribute value in a traceable, reproducible and cost efficient manner. Thus, the decision model refers to an operational level (attribute values) supporting DQM and provides methods to determine the required currency with respect to a particular context. Its operation ability was illustrated using a direct marketing example.

Some limitations of the decision model provide room for further research. Currently, the model is a first step towards quantifying the context-dependent required DQ level as it only refers to the DQ dimension currency. Future research should also consider other DQ dimensions such as completeness, consistency, etc. in order to develop a generalized decision model. Beyond, the dependency between different DQ dimensions should be subject to further research, too. Another shortcoming is the missing empirical evidence of the proposed model. In order to validate the model and its results, several case studies should be conducted.

REFERENCES

1. Askira Gelman, I. (2010) Setting Priorities for Data Accuracy Improvements in Satisficing Decision-making Scenarios: A Guiding Theory, in *Decision Support Systems*, 48, 4, 507-520.

2. Ballou, D. P. and Pazer, H. L. (1995) Designing information systems to optimize the accuracy-currency tradeoff*, Information Systems Research*, 6, 1, 51-72.

3. Ballou, D. P. and Tayi, G. K. (1999) Enhancing Data Quality in Data Warehouse Environments*, Communications of the ACM*, 42, 1, 73-78.

4. Ballou, D. P., Wang, R. Y., Pazer, H. L. and Tayi, G. K. (1998) Modeling Information Manufacturing Systems to Determine Information Product Quality*, Management Science*, 44, 4, 462-484.

5. Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009) Methodologies for data quality assessment and improvement*, ACM Computing Surveys (CSUR)*, 41, 3, 1-52.

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

7. Cappiello, C. and Comuzzi, M. (2009) A UTILITY-BASED MODEL TO DEFINE THE OPTIMAL DATA QUALITY LEVEL IN IT SERVICE OFFERINGS, in *Proceedings of the 17th European Conference on Information Systems (ECIS 09)*, June 2009, Verona, Italy, 1062-1074.

8. 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*, June 2004, Paris, France, 68-73.

9. Even, A. and Kaiser, M. (2009) A Framework for Economics-Driven Assessment of Data Quality Decisions, in *Proceedings of the Fifteenth Americas Conference on Information Systems (AMCIS 09)*, August 2009, San Francisco, California, USA, Paper 436.

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

11. Even, A. and Shankaranarayanan, G. (2009) Dual Assessment of Data Quality in Customer Databases, *ACM Journal of Data and Information Quality*, 1, 3, 15:1-15:29.

12. Fisher, C. W., Chengalur-Smith, I. N. and Ballou, D. P. (2003) The Impact of Experience and Time on the Use of Data Quality Information in Decision Making, *Information Systems Research*, 14, 2, 170-188.

13. Harris Interactive. (2006) Information Workers Beware: Your Business Data Can't Be Trusted, http://www.sap.com/about/newsroom/businessobjects/20060625_005028.epx, 2008-10/13.

14. Heinrich, B., Kaiser, M. and Klier, M. (2007a) DQ metrics: a novel approach to quantify currency and its application in CRM, in *Proceedings of the 12th International Conference on Information Quality (ICIQ 07)*, November 2007, Boston, USA, 431-445.

15. Heinrich, B., Kaiser, M. and Klier, M. (2007b) How to measure data quality? – a metric based approach, *Proceedings of the 28th International Conference on Information Systems (ICIS 07)*, December 2007, Montreal, Canada.

16. Heinrich, B., Kaiser, M. and Klier, M. (2009) A Procedure To Develop Metrics For Currency And Its Application In CRM*, ACM Journal of Data and Information Quality*, 1, 1, 5:1-5:28.

17. Heinrich, B. and Klier, M. (2009) A Novel Data Quality Metric For Currency Considering Supplemental Data, in *Proceedings of the 17th European Conference on Information Systems (ECIS 09)*, June 2009, Verona, Italy, 2701-2713.

18. Heinrich, B. and Klier, M. (2006) Ein Optimierungsansatz für ein fortlaufendes Datenqualitätsmanagement und seine praktische Anwendung bei Kundenkampagnen*, Zeitschrift für Betriebswirtschaft*, 76, 6, 559-587.

19. Helfert, M., Foley, O., Ge, M. and Cappiello, C. (2009) Limitations of Weighted Sum Measures for Information Quality, in *Proceedings of the Fifteenth Americas Conference on Information Systems (AMCIS 2009), August 2009, San Francisco, California*

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

21. Jiang, Z., Sarkar, S., De, P. and Dey, D. (2007) A Framework for Reconciling Attribute Values from Multiple Data Sources*, Management Science*, 53, 12, 1946.

22. 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.

23. Otto, B., Ebner, V. and Hüner, K. M. (2010) Measuring Master Data Quality: Findings from a Case Study, in *Proceedings of the Sixteenth Americas Conference on Information Systems (AMCIS 10)*, August 2010, Lima, Peru, Paper 454.

24. Pierce, E. M. (2004) Assessing data quality with control matrices*, Communications of the ACM*, 47, 2, 82-86.

25. Pipino, L., Lee, Y. W. and Wang, R. Y. (2002) Data Quality Assessment*, Communications of the ACM*, 45, 4, 211-218.

26. Redman, T. C. (1996) Data Quality for the Information Age, Artech House, Boston.

27. Russom, P. (2006) Taking Data Quality to the Enterprise through Data Governance, The Data Warehousing Institute, Seattle, USA.

28. Shankaranarayanan, G., Ziad, M. and Wang, R. Y. (2003) Managing Data Quality in Dynamic Decision Environments: An Information Product Approach*, Journal of Database Management*, 14, 4, 14-32.

29. Umar, A., Karabatis, G., Ness, L., Horowitz, B. and Elmagardmid, A. (1999) Enterprise data quality: A pragmatic approach*, Information Systems Frontiers*, 1, 3, 279-301.

30. Wand, Y. and Wang, R. Y. (1996) Anchoring data quality dimensions in ontological foundations*, Communications of the ACM*, 39, 11, 86-95.

31. Wang, R. Y. (1998) A Product Perspective on Total Data Quality Management*, Communications of the ACM*, 41, 2, 58-65.

32. Wang, R. Y. and Strong, D. M. (1996) Beyond accuracy: what data quality means to data consumers*, Journal of Management Information Systems*, 12, 4, 5-33.