Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2009 Proceedings

Americas Conference on Information Systems (AMCIS)

2009

A Framework for Economics-Driven Assessment of Data Quality Decisions

Adir Even
Ben-Gurion University of the Negev, adireven@bgu.ac.il

Marcus Kaiser
University of Augsburg, marcus.kaiser@wiwi.uni-augsburg.de

A Framework for Economics-Driven Assessment of Data Quality Decisions

Adir Even
Dept. of Industrial Eng. and Mgmt.
Ben-Gurion University of the Negev, Israel
adireven@bgu.ac.il

Marcus Kaiser
Research Center of Finance and Information Mgmt.
University of Augsburg, Germany
marcus.kaiser@wiwi.uni-augsburg.de

ABSTRACT

Economic perspectives have raised growing attention in recent data quality (DQ) literature, as studies have associated DQ decisions with major cost-benefit tradeoffs. Despite the growing interest, DQ research has not yet developed a robust, agreed-upon view for assessing and studying the link between DQ and economic outcome. As a contribution, this study proposes a framework, which links costs to the decisions made in managing the information process and improving the DQ, and benefits to the use of information-product outcomes by data consumers. Considering past research contributions, we develop this framework further into a high-level optimization model that permits quantitative assessment of cost-benefit tradeoffs, towards economically-optimal DQ decisions. We demonstrate a possible use of the proposed framework and the derived model, and highlight their potential contribution to an economics-driven view of DQ issues in both research and practice.

Keywords

Data Quality, Data Management, Metrics, Cost-Benefit Analysis, Information Process, Information Product

INTRODUCTION

Data Quality (DQ) has been studied from different technical, functional, and organizational perspectives. A common theme in DQ studies is that "the higher DQ is – the better". Poor DQ was shown to cause major damages to organizations in terms of operational failures, decreasing trust and reputation and, at the bottom line, profitability loss. Accordingly, studies often treat high DQ as the main objective, and focus on methodologies, tools and techniques for improving it. Indeed, higher DQ has clear merits from many technical, functional and organizational viewpoints. However, in this study we suggest that high DQ should not necessarily be the only objective to consider when evaluating DQ decisions. When economic aspects – such as the benefits gained from high-quality data versus the cost of improvement – are taken into account, higher DQ is not necessarily better. This argument has been supported by recent DQ studies that have shown the criticality of economic aspects in assessing DQ decisions. When targeting high DQ alone as the ultimate end, while ignoring economic aspects, the decision outcome might turn out to be a significant damage to profitability, or even a net loss. This is particularly true with the immense growth in the volumes of data that organizations manage, which implies higher costs due to the need to increase investments in ICT and DQ improvement efforts.

This study argues that DQ decision-making must recognize and consider economic outcomes, as cost-benefit tradeoffs are often significant in that context. Further, it suggests that economics-driven assessment requires establishing a quantitative link between DQ decisions and economic outcomes. As a contribution, this study proposes a framework for assessing economic tradeoffs in a DQ decision process. Following previous research, the framework observes data environments as multi-stage processes with information-product outcomes. The framework links economic benefits (conceptualized as *utility*) to the use of information products, and costs to the production process and DQ improvement efforts. The framework is further developed into a microeconomic model that permits quantitative assessment of DQ decisions, based on maximizing the *net-benefit* – the difference between *utility* and *cost*. We suggest that adopting such a framework will introduce economic thinking into to the design of real-world DQ management processes, and can also guide future research of DQ economics.

We next describe the proposed framework and its different components. While laying out the framework, we also highlight previous DQ studies that had influenced our thinking. Quantitative analysis done in some of those studies also had impact on the development of the microeconomic model that we introduce in the following section. We first introduce the model at a high level, and then demonstrate a possible use of it in a specific DQ decision-making scenario. To conclude, we state the contributions of this study, highlight its limitations, and propose directions for future research.

A FRAMEWORK FOR ECONOMIC ASSESSMENT OF DATA QUALITY DECISIONS

This section lays the foundations for a framework for an economics-driven assessment of DQ decisions (Figure 1). The framework adopts two premises of the Total Data Quality Management (TDQM) approach (Wang, 1998). First is the notion that a data environment can be conceptualized as a complex multi-stage information process, which transforms raw material – data retrieved from different sources – into information products, used by data consumers. Our framework attributes economic benefits to the use of information products, and costs to the implementation and the maintenance of an information process. Second is the argument that DQ improvement is not a one time effort, but rather an ongoing cycle of incremental improvements. The acts of identifying, quantifying, and analyzing economic tradeoffs are linked to the different stages of this cycle, as well as the actions taken toward improving net benefits.

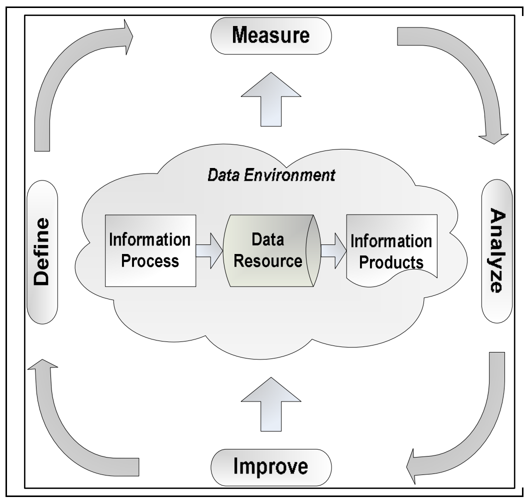


Figure 1. A Framework for Assessing the Economics of Data Quality

Data Resources and Environments

Data resources are often viewed as the "raw material" for information and knowledge. For the purpose of developing our framework, we define a data resource as any structured, electronically-stored collection of data items. A common model for storing structured data resources is the tabular dataset – a two-dimensional model that represent a data collection as multiple records with similar attribute structure. Tabular datasets underlie the popular RDBMS (relational database management system) technology, and Even et al. (2007) show that the design of large tabular datasets may introduce significant cost-benefit tradeoffs. The analytical model that we develop later relates to data stored in a tabular dataset; however, the economic-assessment framework that we introduced in this study is not limited to data stored in tabular datasets.

To manage data resources, organizations invest in data environments that include repositories for storing data resources, applications for processing them, and tools for analyzing and delivering them to data consumers. In these environments, we differentiate between two high-level components, which are described next: *(a) Information Product* – the outcome generated from data resources, used by data consumers; and *(b) Information Process* – a collection of processing stages that feed and update the data resource. From an economic perspective, the former can be associated with the benefits gained from investing in data environments, while the latter can be linked to the DQ cost.

Information Products and Utility

Data resources generate economic benefits through usage by data consumers. The outcomes of data environments, which are generated out of the data resource, have been conceptualized as information products. These information products could be reports, datasets, analyses, or other forms of data integration and presentation. The benefits gained from the use of information products have been conceptualized as utility, which is often (but not necessarily) measured monetarily. Utility contribution of data resources can be gained through their integration into business processes such as tracking stages in the supply chain or managing customer contacts. Benefits can be also gained when data resources are used for business analysis and decision support – in which case, the utility has been conceptualized as the difference in gains between the outcomes of making the decisions without data resources being available, versus the outcomes of making the decision with data-resource support. A third possible form of gaining benefits from information products is using them as a commodity. This is often done by data vendors that specialize in delivering data to other firms.

Information Process and Costs

The multi-stage collection of processes and ICT (information and communication technology) utilities that helps generating and updating the data resource has been conceptualized as an information process (Wang, 1998). Such a process may include stage of data acquisition (e.g., via manual and/or automated processes), processing (integration of data from multiple sources, error cleansing, and/or transformation into usable formats), storage in permanent or temporary databases, retrieval and delivery to consumers. The above activities are associated with costs, as implementing and managing data environments require certain investments, which are often high. Data acquisition costs, for example can be associated with the ICT infrastructure for data collection, and/or labor required for data entry. Data processing solutions may involve purchase of software and associated customization. Storage requires investments in servers and disk space, database management systems (DBMS), as well as database design and administration efforts. Delivery costs may include the purchase of reporting and Business Intelligence (B.I.) tools besides programming and administration expenses.

The Data Quality Improvement Cycle

The TDQM approach suggests that quality management in data environments can be viewed as an ongoing improvement process, containing cycles of *definition, measurement, analysis and improvement* stages (Wang, 1998). We next discuss how these stages can be linked to economic assessment of cost-benefit tradeoff, suggesting that once this link is established – the assessment of economic effects may provide important inputs and insights to DQ decision-making throughout the cycle. For brevity, we describe the framework as addressing quality improvement for a single data resource. This resource resides within a certain data environment that includes an information process for maintaining it, and a set of information products that are generated out of it. Obviously, in many organizational settings we will find multiple data resources, often managed in different but interlinked data environments and the framework ought to be extended in the future accordingly.

Definition

We suggest that the definition stage of the quality improvement cycle has to address four important aspects – the *objective*, the *scope*, the *set of actions*, and a *model* that describes the anticipated effect of these actions. DQ literature commonly defines the key objective of DQ improvement as *fitness to use* – the ability of data resources to satisfy the data consumer's needs. It has been suggested that fitness to use can be assessed along different dimensions – each reflecting a different type of quality defect and/or a different reduction in the ability to use data adequately. Some studies have suggested that the notion of fitness to use must consider the maximization of economic outcomes as an important objective of DQ improvement efforts. Ballou and Pazer (1995, 2003) propose utility as a measure for the benefits gained from data resources. Using utility as an objective, they develop methodologies for assessing DQ improvement tradeoffs. Even et al. (2007) suggest the net-benefit, the difference between utility and cost, as an objective for assessing design decisions in data environments and among these decisions the targeted DQ level. Adopting these previous views, we assume that DQ decisions affect economic outcomes, such as utility and cost, and see the net-benefit as the objective of DQ improvement efforts.

Redman (1996) proposes three possible avenues for improving DQ – each highlights a different *scope: (a) Design:* data environment can be built from scratch, or comprehensively redesigned, to better manage data at a high quality – for example, by embedding controls in processes, supporting quality monitoring with metadata, and improving operational efficiency. A robust design can help eliminate root causes of defects, or greatly reduce their impact. *(b) Process Improvement:* within an existing data environment, improvement efforts may target certain stages of the process, monitor them more closely and reconfigure them to reduce hazardous effects on DQ, and *(c) Error Correction:* quality improvement efforts may focus on the data resource itself – attempting to detect quality defects and use different cleansing and error-correction methods to fix them to an extent. Heinrich and Helfert (2003) see the first two approaches as proactive – they aim at influencing the quality of data to be acquired in the future. In contrast, the latter one can be viewed as reactive – trying to improve the quality of data already stored in the data resource (e.g., with data cleansing actions like correcting data). In this study, we suggest that economic thinking can be applied to both proactive and reactive DQ improvements.

Depending on the scope of DQ improvement, one has to define a set of *DQA – Data Quality Actions*. Certain DQA are proactive in nature; hence are more likely when the scope is designing the data environment, or certain stages within it. If for instance an attribute is added to an existing entity in a database, not only the data resource has to be adapted, but also the information process, as the values for the corresponding attribute must be generated. Moreover, the new attribute values will be used in information products; hence, they will require adaptation as well. Even et al. (2006) associate utility-cost tradeoffs with certain choices of ICT platforms in a data warehouse, which affect the quality of data resources that it manages. Even et al. (2007) associate design decisions in a tabular dataset – such as the time span covered and the set of attributes included – with economic outcomes. In contrast, certain DQA are reactive in nature, as they leave the design unchanged, but rather attempt to improve the quality of the values already stored in the data resource. For instance, Heinrich et al. (2007) seek to improve the benefits gained from a mailing campaign by updating the customers' address data.

Finally, the definition stage has to assess the anticipated effect of each action. When putting this assessment in economic perspective – we propose that each action may affect both utility and cost. Generally, it is reasonable to assume that higher DQ leads to higher utility but, at the same time, also requires greater investments; hence, implies higher cost. We suggest that, at the definition stage, such anticipated effects can be represented in a form of a parameterized model that links DQA to economic outcomes (in the following section we propose a high-level model that can serve this purpose). Such an econometric model directs the evaluation, and can later be enhanced and elaborated, as the evaluation proceeds. The development of such an econometric model requires a definition of *(a) Input,* in the form of a set of design and/or configuration characteristics, which are the subject for decision. *(b) Output*, in terms of economic outcomes such as utility, cost, and net-benefit, and *(c) Effect-Mapping,* a quantitative formulation that links inputs to outputs, towards assessing the possible effect of the different DQA. Such an approach has been applied in a few studies, for instance Ballou and Pazer (1995, 2003) develop utility-driven models for assessing tradeoffs in targeting high DQ level along different dimensions. Heinrich et al. (2007) analyze the relationship between costs, benefits and DQ for the case of a mailing campaign in the field of customer relationship management. Even et al. (2007) develop a decision model for optimizing a tabular dataset, which links design decisions such as attribute selection and the targeted DQ to economic outcomes.

Measurement

To permit the use of the parameterized model for assessing DQA and choosing the optimal among them, certain measurement activities will be required. When considering the targeted quality level as the input to an econometric model, a key question is how to measure it. DQ measurement has been addressed by a plethora of DQ studies, which identified a set of dimensions along which quality can be assessed (e.g., completeness, accuracy, and currency), and proposed different methods for measuring them (e.g., Redman, 1996). Further, studies have differentiated between impartial DQ assessments, driven by rates of defects in the data resource versus contextual assessments, which evaluate quality within a context of use. Even and Shankaranarayanan (2007) link the differentiation between impartial versus contextual quality measurement to the utility of data – suggesting that DQ measurements driven by the presence of defects reflect impartial assessment, while DQ measurements that are driven by the impact of defects on utility are contextual in nature.

Besides measuring the current level of DQ, the measurement stage must also estimate the costs and utilities associated with the DQ level, and the effect of the different DQA on these model outputs. To map the effect, one has to estimate the model's parameters, using appropriate statistical parameter-estimation methods. A plethora of parameter estimation methods has been discussed in literature. Heinrich et al. (2007), for example, assess the quality of the customer address data stored in a data resource by means of a currency metric, determine cost resulting from buying up-to-date addresses from an external provider and derive the utility in terms of expected additional profits. To compute the latter, the authors rely on historical success rates of similar campaigns which are determined according to the currency of customers' address data. In many other cases, determining cost and utility may not be as straightforward as in the given instance. Nevertheless, assessing DQ, costs and utilities is crucial for an economic management of DQ in business environment, as these are essential input factors to the decision model. As estimating economic outcomes might turn out to be complex, further research in this area is required.

Analysis and Implementation

After defining the model and the DQA, and measuring the model's inputs, outputs, and parameters – some analysis will be required to assess the different alternatives and choose the subset of DQA that should be taken. As discussed by Even et al. (2007) an econometric model that maps design and configuration characteristics to economic outcome is an optimization model in nature, in which the optimization objective is to maximize net-benefits (the difference between utility minus cost). We can use a model as such to assess what characteristic values should be targeted, and what DQA should be taken such that economic outcomes are optimized. An optimization model would help evaluating utility-cost tradeoffs between the different DQA from a set of alternatives, and identifying the optimal subset. Obviously, it may turn out, that none of the candidate DQA should be selected, as the resulting benefits from a higher DQ do not justify the associated costs.

Even et al. (2007) provide an illustrative example for a model in which utility and cost grow to a certain power with higher DQ, and show that such a model can be used for optimizing the targeted DQ level. In certain cases, the optimum can be determined by a closed-form solution, while others will require a numerical approximation of the optimum. The example of a mailing campaign in (Heinrich et al. 2007) segments the customers according to the currency of their address data stored in the data resource. Then it is analyzed for which segment updating customer address data provides positive net-benefit, i.e. are the expected additional profits higher than the costs for buying addresses. It turns out, that this holds only for a relatively small fraction of the customers. Besides DQ, cost and utility, the optimization models may need other inputs as well. For instance, the model proposed by Heinrich and Klier (2006) relies on a parameter, which expresses a customer's reaction to DQA. Such parameters are determined using statistical analysis over historical data.

Once the optimal set of DQA has been chosen, the corresponding actions have to be implemented in order to improve DQ and to realize the corresponding benefits. After improvement actions took place, one must verify the extent to which the intended improvement, as predicted by the models, was indeed realized, as the actual DQ level, cost and utility (a posteriori) may differ from the estimated ones (a priori). The differences can then be used to predict the effects of a DQA more precisely in the next iteration of the cycle. For example, after the addresses for the identified customers segments had been bought, Heinrich et al. (2007) examined whether the results of the metrics for currency were a good indicator for the probability, that the customer addresses are up-to-date. It turned out that the percentage of customer addresses predicted to be up-to-date was indeed very close to the actual one.

When entering a cycle of DQ improvement – one must consider the cost involved in implementing the cycle. All cycle stages require time and labor, and may involve some cost – defining the cycle, measuring and estimating the different model components, running the analysis and, obviously, the implementation of the chosen DQA. These costs depend substantially on the improvement's scope – when the scope is large, the cost of implementing the DQ improvement cycle is likely to be high; hence, utility-cost tradeoffs should be assessed more carefully.

AN ECONOMICS-DRIVEN MODEL FOR OPTIMIZING DATA QUALITY

To develop our framework further, we now suggest a high-level microeconomic model, which links DQ design and configuration decisions to economic outcome in a quantitative manner and permits evaluation of alternative DQA towards identifying an economically-optimal action. We then demonstrate an enhancement to the baseline model, which addresses a specific DQ decision scenario.

The Baseline Model

We develop the model for a single iteration of the decision cycle. The input to the model is a vector Q which represents a set of *decision variables* – design and configuration characteristics, which are subject to decision. This set of characteristics reflects the scope of DQ improvement efforts, decided in the beginning of the decision cycle. The model's output is the added net benefit B, which becomes the objective for optimization. The added net-benefit has the following components:

- *U:* Improvement to the utility gained by the use of information products.
- *C:* Cost that was added (or saved) in the information process. We assume that some reduction to utility can be tolerated, if the cost that was saved was significantly higher.
- *D, M, A, I:* costs attributed to the different stages of the decision process – *Definition*, *Measurement*, *Analysis*, and *Implementation*, respectively.

We assume that each of these components may be affected to an extent by the decisions made. We hence, represent these components as functions that map the decisions made (the setting of Q) to monetary value. Accordingly, the net benefit can be also represented as a function of the decision variable Q: $B(Q)=U(Q)-C(Q)-D(Q)-M(Q)-A(Q)-I(Q)$. In certain cases, it is possible that the some costs do not depend on the decision made – hence, can be represented as a constant.

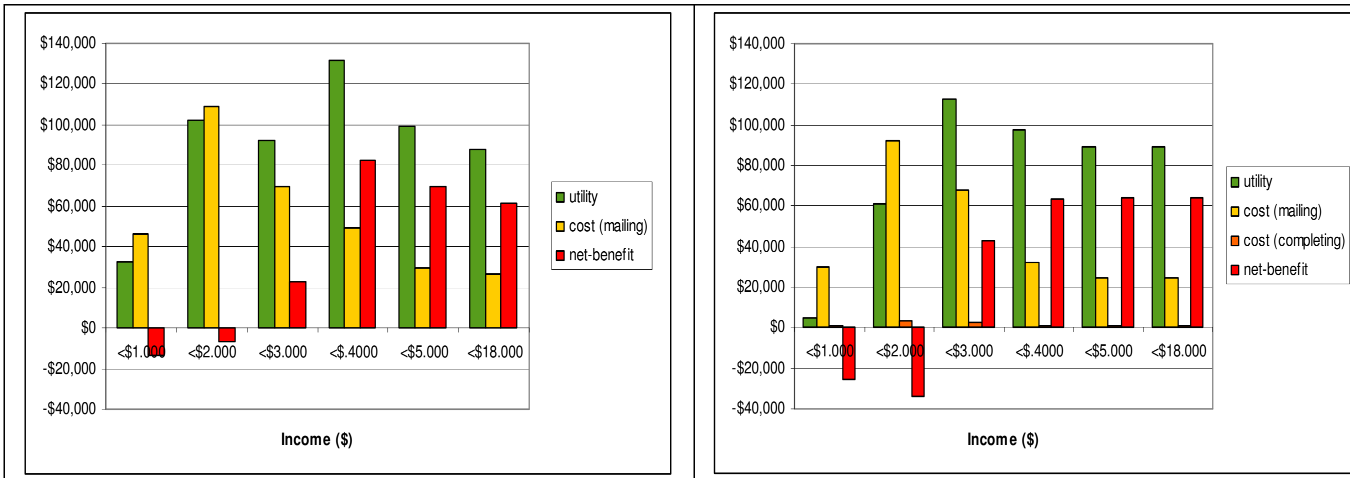
We now define a set of DQA – a total of J actions, indexed by $[j]$. Each action presents an alternative plan for DQ improvement and the assumption is that for each plan, an optimal level of the decision vector Q^*_j will be chosen, such that the net benefit $B(Q^*_j)$ is maximized.

To summarize, we can formulate the decision problem as – choose the optimal DQA $[j]$, and the corresponding optimal level of the decision vector Q^*_j, such that the added net-benefit $B(Q^*_j)=U(Q^*_j)-C(Q^*_j)-D(Q^*_j)-M(Q^*_j)-A(Q^*_j)-I(Q^*_j)$ is maximized.

Enhancing the Model to Address a DQ Decision Scenario – an Illustrative Example

Next, we illustrate an economic assessment of a DQ decision, using the model. Let us assume a financial-service provider (FSP) plans to conduct a mailing campaign for offering private health insurance. In a former, similar campaign, the FSP addressed all of its 200,000 customers, resulting in a success rate of 5%. For each customer that accepts the offer, the FSP gets a commission of $100; hence, the overall utility is estimated at $1,000,000. The mailing costs were $3 per customer – hence, a total mailing cost of $600,000 and a net-benefit of $400,000.

In the *definition* stage, it has been suggested to use the 'Income' attribute when selecting the customers that will be targeted in the upcoming campaign, where the objective is to improve performance in terms of net-benefit. A review of the former campaign results indicated that the higher is the income, the higher was the likelihood of a customer to accept the offer. Income is known only for 110,000 customers (55% of the dataset), and for those customers, the FSP determined utilities, mailing costs and net-benefit, per income level (Figure 2a).



(a) (b)

Figure 2. Campaign results for customers with (a) complete versus (b) incomplete 'income' data

As income data is not available for 90,000 customers (45% of the dataset), the FSP has decided that the scope of DQ improvement efforts will be the correction of this deficiency. The identified *set of actions*, which are reactive in nature, consists of: (1) paying for a study, which can provide an estimation of income figures for a given age and occupation (both attributes are stored in the database); and (2) Paying an agency that will verify the income level for each customer. The general assumption is that, in general, a higher targeted quality level of the ‘income’ attributes will improve utility but, at the same time, will also increase the cost.

The *measurement* stage will focus on the 'completeness' metrics, which reflect the extent to which attribute values are missing. As 110,000 out of 200,000 customers' records contain a value for the attribute 'income', the current completeness level is 0.55. The study costs $10,000, whereas the agency charges $5 per customer. We assume that taking either action will complete each of the 90,000 customers' income data ($Q^*_j = 1$). Hence, we can assign the costs for the study to $0.11 per customer. As the added utility of both actions is the same, the first action (the study) is preferable ($j^*=1$). For this action, we can compute the added net-benefit for completing the income data for the 90,000 customers, taking into account the utility as well as the costs for mailing and completing the data, based on the success rates of the former campaign.

In the *analysis* stage, we consider the results of the previous campaign (Figures 2a and 2b) and see that the utility was offset by the mailing costs for the lower income classes (<$2,000). This is due to the fact that a major part of the customers belonged to these lower-income classes and, in addition, these classes had lower success rates. Therefore, the FSP should possibly consider avoiding these customers in the forthcoming mailing campaign. Had the FSP not targeted these customers in the former campaign – the net-benefit would have been $236,500 for the customers with a complete 'income' data and $233,900 for the rest. That means, the former campaign would have provided a net-benefit of $B(1)=$$470,400, about $70,000 higher than the realized one $B(0.55)=$$400,000. Hence, the FSP determines $Q^*_1=1$ and addresses only those customers with a positive net-benefit, i.e. whose income is >$2,000. Assuming that the total cost for the DQ improvement stages ($D(Q^*_1)$, $M(Q^*_1)$, $A(Q^*_1)$, $I(Q^*_1)$) does not exceed net-benefit $B(1)$, the FSP takes action j=1.

In the *implementation* stage, the FSP completes the data resource, using the suggested study. The results of the forthcoming campaign concerning success rates can be used to verify whether the anticipated results were indeed achieved. The example illustrates the fact that measuring and improving DQ do not necessarily improve economic advantage on their own. However, they can serve as the baseline for making decisions towards gaining a higher net-benefit.

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

In this paper, we highlight the need for a robust economic thinking in DQ management. We propose a framework for economics-driven assessment of DQ decisions and, based on this framework, we develop a quantitative high-level model that links DQ decision to economic outcomes. An illustrative example extends the model to address a DQ-decision in customer management environment, which highlights the effects of taking into account economic considerations. As demonstrated with the example, the framework and the model proposed may help practitioners evaluate economic tradeoffs in real-life DQ management scenarios. We also believe that the proposed framework may help directing research on economic effects and tradeoffs in DQ management and decision-making.

The framework proposed in this study and the quantitative model are at a preliminary stage and will require some more development and enhancements before they can be used in research and practice. The illustrative example analyzes a certain type of decision, regarding the improvement of completeness. As highlighted by previous DQ research – the question of data quality is a lot more complex, and it can be observed along many different dimension of analysis; hence, the framework should be extended to a set of different models, each addressing different quality aspects. A fuller model should also address multi-periodic decision scenarios – sequences of DQ decision cycles in which the outcome of one affects the following. Such an extension is required, as many DQA cannot reveal their full value immediately upon implementation and required follow-up DQA. Further, certain DQA enhance the data resource in a way that create opportunities for developing new forms of usage, hence, increase utility – what brings into mind the use of real-option modeling.

While suggesting that economic thinking has important merits to DQ management, we do not see it as replacing other important DQ perspectives, but rather as complementing. As data environments are complex and rapidly changing, it requires developing a broad perspective, which takes into account technical, functional, and organizational aspects, and in addition, as we suggest in this study – also economic ones.

REFERENCES

Ballou D.P., and Pazer H.L. "Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff," Information Systems Research (6:1), 1995, pp. 51-72

Ballou D.P., and Pazer H.L. "Modeling Completeness versus Consistency Tradeoffs in Information Decision Systems," IEEE Transactions on Knowledge and Data Engineering (15:1), 2003, pp. 240-243

Even, A., Shankaranarayanan, G. and Berger, P. D. (2006), "The Data-Warehouse as a Dynamic Capability: Utility-Cost Foundations and Their Implications for Economics-Driven Design", The Intl. Conf. on Information Systems (ICIS), Dec. 2006, Milwaukee, WI

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

Even, A., Shankaranarayanan, G. and Berger, P. D. (2007) Economics-Driven Data Management: An Application to the Design of Tabular Datasets, IEEE Transactions on Knowledge and Data Engineering, 19, 6, 818-831.

Heinrich, B. and Klier, M. (2006) A model to optimize continuous data quality management and its practical application for customer campaigns (in German), Journal of Business Economics, 76, 6, 559-587.

Heinrich, B., Kaiser, M. and Klier, M. (2007) How to measure data quality? – a metric based approach, in Suzanne Rivard and Jane Webster (Eds.) Proceedings of the 28th International Conference on Information Systems (ICIS), December 2007, Montreal, (Canada).

Helfert, M. and Heinrich, B. (2003) Analyzing Data Quality Investments in CRM - A model-based approach, in Martin J. Eppler, Markus Helfert, Leo Pipino and Arie Segev (Eds.) Proceedings of the 8th International Conference on Information Quality, November 2003, Cambridge/Boston, (USA), 80-95.

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

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