Measuring Master Data Quality: Findings from a Case Study

Boris Otto  
*University of St. Gallen*, boris.otto@unisg.ch

Verena Ebner  
*University of St. Gallen*, verena.ebner@unisg.ch

Kai M. Hüner  
*University of St. Gallen*, kai.huener@unisg.ch

Follow this and additional works at: [http://aisel.aisnet.org/amcis2010](http://aisel.aisnet.org/amcis2010)

Recommended Citation  
[http://aisel.aisnet.org/amcis2010/454](http://aisel.aisnet.org/amcis2010/454)

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Measuring Master Data Quality: Findings from a Case Study

Boris Otto  
University of St. Gallen  
Institute of Information Management  
boris.otto@unisg.ch

Verena Ebner  
University of St. Gallen  
Institute of Information Management  
verena.ebner@unisg.ch

Kai M. Hüner  
University of St. Gallen  
Institute of Information Management  
kai.huener@unisg.ch

ABSTRACT
Data quality management plays a critical role in all kinds of organizations. Data is one of the most important criteria for strategic business decisions within organizations and the foundation for the execution of business processes. For the assessment of a company’s data quality, to ensure the process execution and to monitor the effectiveness of data quality initiatives, data quality has to be measured in the same way over a certain period of time. This can be achieved by implementing a measurement system. By now, the implementation of such a system to measure data quality is realized in very few organizations. This paper presents a case study describing the implementation process of a Master Data Quality Cockpit as well as the system used for measuring. The study assesses organizational, process-related, and system level changes as well as success factors necessary to implement such a tool.

Keywords
Data Quality Management, Data Quality measures, Master Data Management, Master Data Quality.

INTRODUCTION
Motivation and Problem Statement
Data quality management (DQM) plays a critical role in organizations (Pipino, Lee and Wang, 2002). Data is the foundation for a number of strategic business requirements such as a 360° view on the customer, efficient and effective decision-making and supply chain integration. However, nearly every IT system has “dirty” (erroneous) data. About 75% of organizations have identified costs originating from defective data (Marsh, 2005). U.S. businesses pay $600 billion a year due to a lack of data quality (DQ) (Eckerson, 2002).

To improve DQ as well as to evaluate the current status, the improvement of DQ and the effect of DQ initiatives (i.e. initiatives for improving DQ or for lowering the risk of data defects) have to be measured. Several authors point out that: “Only what can be measured can be improved” (English, 1999; Wand and Wang, 1996; Wang and Strong, 1996). What is needed is a measurement approach to continuously determine the level of DQ and its impact on the business over time.

Research Question and Paper Structure
The research objective of this paper identifies success factors for implementing a measurement system to measure DQ from a successful case. This objective addresses the following questions:

• What are requirements for DQ metrics on a strategic, organizational and system layer?
• How can DQ metrics be identified?
• How can these metrics be implemented in a tool?
• How does the regular operations process / monitoring look like?

The research paper presents the case study at a chemical company. It focuses on the approach to implement the measurement system and outlines the actions and details necessary for a successful launch. This study does not include a generalized approach to establish DQ measures and only gives the view of the company it was realized at. The proposed tasks do not guarantee completeness.

The paper is structured as follows. The following section gives an overview of the background information, especially the terminology of data, DQ and master data management (MDM). Section three describes the research approach and its limitations. In section four, the project is presented. The paper closes with a summary and an outlook on further research objectives and activities planned in section five.

BACKGROUND

Master Data and Master Data Management

Information systems provide data in a certain business context. When data is being used by human beings, they turn into information, and information finally turns into knowledge by interpreting and linking information for a given purpose (Boisot and Canals, 2004; Davenport and Prusak, 1998; Spiegler, 2000; Stahlknecht and Hasenkamp, 2005).

Master data stores and describes features of a company’s core entities, which most notably are customers, suppliers, products, materials, and employees (DAMA, 2008; Dreibelbis, Hechler, Milman, Oberhofer, van Run and Wolfson, 2008; Loshin, 2008). Typically, master data is used across multiple business processes (e.g. supplier master data is used both by procurement departments and by accounting departments) and is often stored in and/or used by multiple application systems.

Master data can be distinguished from other data, such as transaction data or inventory data, using the following criteria (Dreibelbis et al., 2008; Mertens, 2000; White, Newman, Logan and Radcliffe, 2006): time reference, modification frequency, volume stability and existential independence.

The master data regarded in the organizations varies between industries as well as between organizations. But there are commonalities between the core classes on the basis of which a classification can be applied (Loshin, 2008). Dreibelbis et al. (2008) identify three categories: product, party and account addressing the questions “Who?”, “What?” and “How?”.

MDM aims at creating an unambiguous understanding of a company’s core entities (Smith and McKeen, 2003). As an application-independent process, it ensures the consistency and accuracy of these data by providing a consistent understanding and trust of master data entities supported by mechanisms to use consistent master data throughout the organization and managing the change. These goals are achieved by implementing a corporate framework including the domains organization, processes and architecture (Dreibelbis et al., 2008).

Data Quality and Data Quality Management

Data is defined of high quality if it has the ability to satisfy the requirements for its intended use in a specific situation. This is often referred to as “fitness for use” (English, 1999; Olson, 2003; Redman, 2001). The intended use is commonly described as a multi-dimensional construct consisting of a set of quality attributes, called DQ dimensions which are determined by the data consumer (Wang and Strong, 1996). On the one hand, DQ plays an important role in the success of a MDM. It supports the trustworthiness of master data and its techniques can be applied in the implementation process. On the other hand, MDM can improve DQ. It reduces the error rate as there is a high probability of identifying mistakes by the integration of multiple systems (Loshin, 2008).

RESEARCH APPROACH

The research method applied within this paper is the single case study research (Yin, 2002). This method was selected due to the little knowledge available within the field of measuring DQ. Within a study (cf. Otto and Ebner, 2010), an expert survey was conducted among organizations, to identify the importance and progress of DQ measurement. The results showed that there are only few organizations already measuring the quality of their data and even less having a measurement system implemented. The case presented within this document shows the approach of a company having already successfully established such a system.
An inductive research approach was chosen to understand the problem at hand and develop an approach to solve it, by identifying best practices and generalizing the actions applied. Within this paper, we present an exploratory single case study. A single case is taken due to the little expertise of companies within this topic. The case study allows to analyze the implementation of a system for measuring DQ within an organizational context. This is necessary, as the measurement equally concerns the organization’s strategy, its processes as well as its technical systems.

The data collection was executed by expert interviews within the selected company. Table 1 shows details to the interviews, the interview partners and their responsibly during the project. The interviews were recorded by protocol and afterwards transcribed, analyzed and evaluated. The approach follows the guidelines for case study research in the field of business engineering proposed by (Senger and Österle, 2002).

<table>
<thead>
<tr>
<th>Interview</th>
<th>Job role</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview 1</td>
<td>Responsible for MDM and DQM within the central DQ unit.</td>
<td>90 min.</td>
</tr>
<tr>
<td></td>
<td>Project leader for MDM projects within the central DQ unit.</td>
<td></td>
</tr>
<tr>
<td>Interview 2</td>
<td>Project leader for MDM projects within the central DQ unit.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical leader of the Data Quality Cockpit.</td>
<td>180 min.</td>
</tr>
<tr>
<td></td>
<td>Project leader for MDM projects within the region Asia-Pacific.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Interview details.

**CASE PRESENTATION**

**Company Profile**

The case was recorded at one business division of a large chemical group within Europe, in the following called ABS. ABS was founded through the merger of two organizations and has several areas of business operations. ABS is market leader in its product area. It has about 20,000 employees and a revenue of approximately 6.5 billion Euro in 2008.

ABS is acting on a global market which is strongly regulated. Its compliance requirements are retrieved from regulations, industry-specific policies and governmental specifications, which derive from the characteristics of the manufactured products and legal requirements of the particular countries, the products are sold in. For example, each product has to be registered and approved before its distribution on the market. The licenses have to be applied for in each country independently. Within a license, the composition of the product and the period of time it is valid in, are defined. Selling a product on the market implies a valid, unexpired license and the conformity of its substance. For this reason, MDM at ABS has to guarantee the generation and allocation of the data and its DQ, necessary for the concession. This avoids losses in revenue, fines and the withdrawal of licenses. It leads to the need that the product master data has to contain data that identifies the countries, the license exists in. Other business drivers within the chemical industry include high costs for research and development as well as long cycle times for research.

**Initial Situation**

Having a background of mergers and acquisitions, ABS has a very heterogeneous application architecture. Due to this fact, an universal strategy to consolidate the heterogeneous infrastructure was installed and different projects were launched to facilitate this strategy. The first project covered the harmonization of the systems and processes within the regions Europe, Asia-Pacific, and the Americas followed by a project to consolidate these three regional systems and processes to a global one. Within these projects, as well as another project to implement a data governance organization, the need for DQM as a management function was observed. Figure 1 shows the progression from the previous landscape to the target landscape after the global consolidation.
Within the corporate strategy no measurement and monitoring of DQ was considered. A measurement of process performance existed and was conducted by validation checks within the applications. The measurements consisted of 25 rules concerning technical correctness and plausibility of data, i.e. whether a value is within a given range or if two fields are consistent. Rules applicable for all regional companies were integrated within the applications. The ownership of the data resided within the countries, but there were no agreements by contract with the employees or a service provider to guarantee targets with respect to the DQ.

For the first consolidation project to migrate to the three regional systems the IBM Information Server, further called IS, was utilized to support the extract, transform and load functionalities, as well as data-cleansing activities. During the selection of the migration tool, capabilities for measuring DQ were taken into account, although no concrete measurement was planned and no budget was drawn up for this part.

During the migration project, the need for measuring DQ arose. On one hand, it was not possible to assess the effect of the projects on DQ and the progress which were achieved. On the other hand, the improvement of DQ was not visible. The idea of measuring DQ came from the central DQ team, which tried to generate interest within the operational departments in the regions. The final initiation for implementing DQ measures took place when the management of Asia-Pacific committed its support during a kick-off workshop of the global harmonization project. This workshop in 2008 concerned the process harmonization of the financial and production planning processes in Asia-Pacific and identified problems in the complex planning process due to erroneous and incomprehensible data.

**Master Data Quality Cockpit**

**Targets and Scope**

To reach the overall objective of improving business process performance by enhancing DQ, the following targets were identified for the implementation of DQ measures.

- Get a statement for the success of the harmonization projects.
- Aim at establishing a healthy competition between the countries concerning the DQ.
- Identify needs for action and the necessity for data management.
- Establish an extensible infrastructure for including future requirements.

The measurement considered only a qualitative rating. A monetary rating of the DQ was not ascertainable and not intended.

To achieve the first two targets, the project was supposed to identify and implement a single, comparable value to measure DQ, the Data Quality Index (DQI). The third target should be reached by establishing a target DQI. A DQI of 99% was the value that had to be reached to complete the migration project. After the completion of this project, a value of 97% was chosen that had to be reached by the regional data owners. The value of 97% was chosen arbitrarily.
Regarding the cost-efficiency ratio between a higher DQ and its benefit, a target DQI of 99% seems to be not effective. Due to the fact that the value of the DQI is established by using business rules, actively being responsible for defects in critical business processes, the DQI has to cover 100%. Choosing a lower value would not ensure the necessary adherence to the critical business processes.

The regional scope of the project was Asia-Pacific comprising 15 countries. The measuring frequency should be regularly once per month. With a monthly occurrence, the coordination of necessary tasks is rather uncomplicated and the effort to eliminate errors is comparatively low. The project focused on the master data material master data for the planning processes in Asia-Pacific.

**Implementation of the Master Data Quality Cockpit**

The implementation of the measurement tool was launched without prior planning or budgeting. The execution was conducted in the following steps. Starting with preliminary work for the project, the measurements were designed, tested and implemented. Finally, after reaching the project targets, the regular operation is established.

As preliminary work, the awareness was built within the operational business units by showing the negative impact of defective data on the business. These relations were illustrated as so-called “causal chains”, where one chain comprises a business problem and data defects, potentially causing this problem. For example, when the profit center is not consistent with the product hierarchy there is an impact on the profitability analysis, cash-flow evaluation, asset accounting and PL accounting. This leads to the risk of wrong decisions concerning the product portfolio revision.

As starting action, some organizational changes were taken to establish tasks and responsibilities for DQ. To improve the business process performance and to support the execution of the activities necessary, one person accountable had to be identified. To emphasize the importance of the initiative the targets were assigned to the personal objectives. BCS announced the country lead of Asia-Pacific to be this person in charge and assigned the DQI of 97% to his target system. He also was supposed to install a master data coordinator who has the task to manage the monitoring of the DQI and to take actions for the rectification of DQ.

As a basis for the DQI, business rules were identified by interviews with the process experts. Beginning with the material-based planning data, the supply chain management experts were consulted to identify the required business rules. In doing so, 160 rules were determined.

<table>
<thead>
<tr>
<th>Rule ID</th>
<th>Business Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>XOM_0038</td>
<td>Profit Center is not consistent with Product hierarchy.</td>
</tr>
<tr>
<td>XPM_0005</td>
<td>Procurement type “E” and valuation class “3xxx”.</td>
</tr>
</tbody>
</table>

Table 2. Examples of validation rules.

For further processing, the rules were tested with regard to three characteristics.

- First, the measurability of the rule, i.e. the rule is technically computable with the data available.
- Second, the impact of the business rule on the business processes. This proofs the causal relation between the rule and the key performance indicators (KPI) of the processes.
- Third, the adherence to the rule within the processes, i.e. a violation is an exceptional defect in the process, and ensures the comparability of the number of defective records between the business rules.

After validating these three characteristics, 53 rules remained. In multiple iterations, the rules were verified and corrected with the support of the process experts until the definition corresponded to the operational behavior. Table 2 shows examples of these rules.

For the identification of the implications of the business rules they were organized in seven groups like costing, material status, packaging, etc. Each group was assigned to a legal entity, e.g. India or Japan. The legal entities in turn were associated to one of the three regions, Europe, Americas and Asia-Pacific. By aggregating all those rules to the top level, a single DQI was constructed, which follows the subsequent definition:

\[
\text{DQI} = 100 - \frac{\text{defective data records} \times 100}{\text{total relevant data records}}
\]
Simultaneous to the actions aforementioned, the technical components were set up. To perform the extraction as well as the validation of the rules IBM Information Server (IS) was used. It was already budgeted within the migration project. For displaying the results within web-based reports, Oracle Application Express (APEX) was used as standard within the organization. The relevant data were extracted from the master data system and loaded into the IS. In foresight to future requirements, not only the material data was integrated into the load process, but also the customer and supplier data. The tables were loaded one-to-one from the source to the target system to identify and track violations against the rules.

The implementation of the business rules in the IS was the next step of the project. For storing the results of the business rule validation a database is used. These values are extracted and processed for the presentation on web-based reports. The implementation of the user interface was realized with APEX. The introduction of a data warehouse for the processing of the data did not seem to be necessary.

Project Status
Currently, the master data objects of the classes material, customer and supplier data are available at the IS for all regions and countries. The business rules identified for the region Asia-Pacific are implemented and executed for all regions on a monthly basis. For the master data classes customer and supplier, the identification of the business rules is in progress.

The organizational foundation is only implemented and in operation in Asia-Pacific, were the awareness for the impact of DQ on the business processes exists and the DQI is integrated into the incentive system. Throughout the project, it was possible to foster the confidence of the business experts by being supported by the central DQ department. Now, the demand for DQ increases due to the positive effect on process performance which was realized by the business departments during the project. In all other regions, the DQI is not integrated into the personal objectives, but the awareness of DQ grows.

As next steps, an expansion of the organizational structure within the regions Americas and Europe is planned. An extension of the master data classes to customer and supplier data started in March 2010. The integration of transactional data, like SAP ERP, for compliance reasons is considered.

Regarding the technical realization, the integration to a data warehouse solution is considered for the near future.

Problems and Learnings
To launch an initiative like this, lots of activities have to be performed to gain the attention and support needed for its execution. For this reason, it is helpful to launch such a measurement initiative in parallel to an existing project, contributing to the success of that other project. This makes it easier to get the support required and also reduces the costs, especially when necessary tools are already available and don’t need to be budgeted separately.

Another problem identified, is the necessity of awareness and support both from management as well as the process experts and users, involved in the execution. Therefore, it is important to demonstrate the impact DQ has on the business processes and which benefits can be achieved by improving DQ. These benefits have to be stated through process improvement and compliance. A monetary rating was not identifiable within the project.

For the definition of business rules and for the conduction of DQ initiatives, collaboration with the operational business department is essential. To get the initial support of these departments at the beginning of the project, incentives have to be provided by integrating the project targets into the personal objectives of the employees. In the course of the execution, the motivation will rise by establishing a healthy competition between the participating countries. In the further progress of the project, the process experts will start to recognize the benefits of increased DQ, like improvements in process efficiency.

For the DQM tasks are new to the process experts, assistance with the operational execution is needed. This need can be satisfied by installing a support team, available to help with problems and questions. It also shows the need for the collaboration of the operational business departments and the information technology department, to combine the knowledge of both and strengthen the success of the initiative.

For being able to demonstrate the tool as early as possible to the business users to get their trust and confidence, an iterative implementation approach is very helpful be used. The refinement can take place in subsequent iterations.

New Solution
The ownership of the data resides with the regional business process experts. The country leads of Asia-Pacific are responsible for the execution of DQ initiatives and have the DQI integrated within their personal objectives. So the master data coordinator, announced by the country leader is responsible for the DQI and for managing DQ within his country.
The ownership and responsibility for the functional capability and technical availability of the Data Quality Cockpit lies at the central DQ department. The tools are owned on behalf of the business. The DQ department also offers guidance with problems and questions on the Data Quality Cockpit, DQM and the execution of necessary actions. This service is provided by the department without internal charges and is not planned to be priced. The regional business department is in charge of managing the DQI, but the DQ department also informs them when an incident is detected by them. There are no automatic notifications or warnings implemented when the DQI exceeds a given threshold.

Changes or the definition of new business rules is initiated by the process experts within the country, if local rules are affected. Every change or renewal of a business rule runs through the initial definition process as stated in section Process Execution. Due to the correction and validation of the rule, a change takes three to six months until going productive. The business rules also support versioning.

The master data of the source systems is loaded monthly into the IS. Within the IS, the business rules are executed after loading and the values measured are stored in the database. Archived measurement values date back as far as 15 months. This retention period was not chosen empirically, five quarters just seemed reasonable and a comparison to the quarter of the preceding year is possible. The reports are presented on the user-interface and provide aggregation levels corresponding to the business rule groups mentioned in the preceding section. The reports were realized with APEX.

The values shown on the reports are the target DQI, the DQI of the own country, the DQI of the best country, mentioned by name, and the DQI of the worst country, not mentioned by name. The aggregation levels within the Cockpit are: region, legal entity, business rule group and business rule. Changes in the DQI can be drilled down to the object violating the business rule. For each rule, a documentation page is available, containing a description of the rule, its impact on the business process and tasks to be performed for correction. A contact person for further questions is also mentioned.

The access to the Data Quality Cockpit is not restricted. Every employee can get access to the cockpit and is able to view the same reports as everyone else. There are only some non DQ reports which are restricted to a small group of users. Each user can define an individual entry screen on the cockpit.

Project Results and Findings

The key findings identified within the project are the following success factors:

- Simplicity: A single, simple metric was considered more meaningful than an array of multiple, detailed metrics and statistics.
- Comprehensibility: It was very important to clearly illustrate the impact on business process to the user. The impact should be easy to understand.
- User-friendliness: The presentation of the report was supposed to be easy to use, with a design addressing the target audience and inclusion of graphics.
- Technical support: The support of the users by a central DQ team was regarded as very important to show the users how the tool supports their daily business.
- Functional support: The operational business departments were considered as the most important resource of domain knowledge for the definition of the business rules and the execution of DQ initiatives.
- Management attention: The sponsorship of the management is seen as key for DQM. By establishing strategic and organizational structures the basis for the initiatives is set up.
- User-acceptance: Motivation and collaboration of the users is supporting the execution of the initiative.
- People support: During the project it became apparent that the management as well the user support is an essential need.
- Collaboration: The operational business departments and the IT department have to work hand in hand for a successful execution.

SUMMARY AND OUTLOOK

For the initiation of a project to implement a Data Quality Cockpit, the study shows the importance, not to emphasize on the monetary rating of the initiative like cost savings. The effects achieved have to be stated with respect to process performance and compliance aspects. Another important fact is the collaboration of IT and business department. A project like this needs the input of the business operations to identify the effects on the business. As well, the technical knowledge and guidance concerning DQ is necessary to conduct those tasks.
The case study also depicts that the successful implementation of a system to measure master DQ is supported by a corporate approach, which combines the organizations strategy, processes and systems. For there are only a few organizations having already implemented a measurement system, it is important to identify the essential tasks to successfully introduce such a system.

Against the background of the study at hand, future research will focus on gaining more insight in projects on measuring master DQ and validate the findings from this study. This is planned to be conducted by further case studies. Those studies will build the foundation for identifying best practices and a generalization of the results. A framework containing all necessary tasks to implement a measurement system will be the outcome of this research-in-progress. The framework should assess the complete extent of DQ problems from its root causes to its impacts. A sample evaluation as well as reconciliation with the target group is planned to verify the findings.

REFERENCES