Barriers to master data quality

Anders Haug and Jan Stentoft Arlbjørn
Department of Entrepreneurship and Relationship Management, University of Southern Denmark, Kolding, Denmark

Received September 2010
Revised November 2010
December 2010
Accepted December 2010

Abstract

Purpose – While few would disagree that high data quality is a precondition for the efficiency of a company, this remains an area to which many companies do not give adequate attention. Thus, this paper aims to identify which are the most important barriers preventing companies from achieving high data quality. By improving awareness of barriers on which to concentrate, companies are put in a better position to achieve high quality data.

Design/methodology/approach – First, a literature review of data quality and data quality barriers is carried out. Based on this literature review, the paper identifies a set of overall barriers to ensuring high data quality. The significance of these barriers is investigated by a questionnaire study, which includes responses from 90 Danish companies. Because of the fundamental difference between master data and transaction data, the questionnaire is limited to focusing only on master data.

Findings – The results of the survey indicate that a lack of delegation of responsibilities for maintaining master data is the single aspect which has the largest impact on master data quality. Also, the survey shows that the vast majority of the companies believe that poor master data quality does have significant negative effects.

Research limitations/implications – The contributions of this paper represent a step towards an improved understanding of how to increase the level of master data quality in companies. This knowledge may have a positive impact on the data quality in companies. However, since the study presented in this paper appears to be the first of its kind, the conclusions drawn need further investigation by other research studies in the future.

Practical implications – This paper identifies the main barriers for ensuring high master data quality and investigates which of these factors are the most important. By focusing on these barriers, companies will have better chances of increasing their data quality.

Originality/value – The study presented in this paper appears to be the first of its kind, and it represents an important step towards understanding better why companies find it difficult to achieve satisfactory data quality levels.

Keywords Data handling, Quality management, Information media

Paper type Research paper

1. Introduction

Almost all activities in organizations involve the use of data, which is the foundation for operational, tactical and strategic decisions. Therefore, if a company aims for efficiency, it is critically important that the data of the company is of adequate quality, i.e. high-quality data is crucial to a company's success (Madnick *et al.*, 2004). Unfortunately, many industry expert surveys indicate that data quality is an area which many companies seem to overlook or at least not give sufficient attention (Marsh, 2005). This paper aims to identify which are the major barriers preventing companies from achieving high data quality. Being aware of which barriers to focus on, companies are provided with a much stronger basis for achieving high quality data. Such insights may also be of great importance for research focusing on methods and techniques for data quality improvement.

Journal of Enterprise Information Management
Vol. 24 No. 3, 2011
pp. 288-303

1741-0398
DOI 10.1108/17410391111122862

Vayghan *et al.* (2007) classify the data that most enterprises deal with into three categories: master data, transactional data, and historical data. Master data is defined as the basic characteristics of instances of business entities such as customers, products, employees, and suppliers. Typically, master data is created once, used many times, and does not change frequently (Knolmayer and Röthlin, 2006). On the other hand, transaction data describes relevant events in a company, e.g. orders, invoices, payments, deliveries, storage records, etc. Since transactions use master data, if the master data is not correct, the transactions do not fulfil their intended purpose. Errors in master data can have significant costs; for instance, if the address of a customer is wrong, this may result in orders and bills sent to the wrong address; if the price of a product is wrong, the product may be sold below the intended price; if a debtor account number is wrong, an invoice might not be paid in time; and so on. Therefore, even a slight amount of incorrect master data can absorb a great part of the revenue of a company. In this context, Knolmayer and Röthlin (2006) argue that capturing and processing master data are error-prone activities wherein inappropriate information system architectures, insufficient coordination with business processes, inadequate software implementations, or inattentive user behaviour may lead to disparate master data.

The data quality problems that many companies face today may be related to technological development in the last decades. The development of information technology has enabled organizations to collect and store more data than has ever before been possible. But, as the data volumes increase, so does the complexity of managing it. Thus, it has been argued that the risk of poor information quality has increased, since larger and more complex information resources are being collected and managed in organizations today (Watts *et al.*, 2009).

As later revealed by the literature review of this paper, academic research based on empirical evidence from multiple companies in relation to data quality is rather sparse despite the importance of having correct and adequate data in a company. However, since there seems to be general agreement that the lack of data quality is a great problem in many companies, this paper seeks to provide empirical insight into why this is the case. More specifically, the paper focuses on identifying the main barriers for achieving high data quality in industrial companies. Since, the ways in which master data and transaction data are created and used are fundamentally different, this paper focuses only on master data in the empirical part, in order to be able to obtain more exact answers from the respondents of a questionnaire study. Thus, the research question this paper addresses is, "What are the main barriers for achieving high master data quality in industrial companies?"

The remainder of the paper is structured as follows. First, the research methodology of the paper is described. Next, Section 3 reviews and organises literature on data quality. Section 4 presents and discusses the findings from the questionnaire survey. The paper ends with a conclusion in Section 5.

2. Research methodology

The paper consists of two main steps in relation to investigating the question in focus. The first step is to review and organise literature, while the second step is a questionnaire study. The research methodology is shown in Figure 1 and subsequently further described.

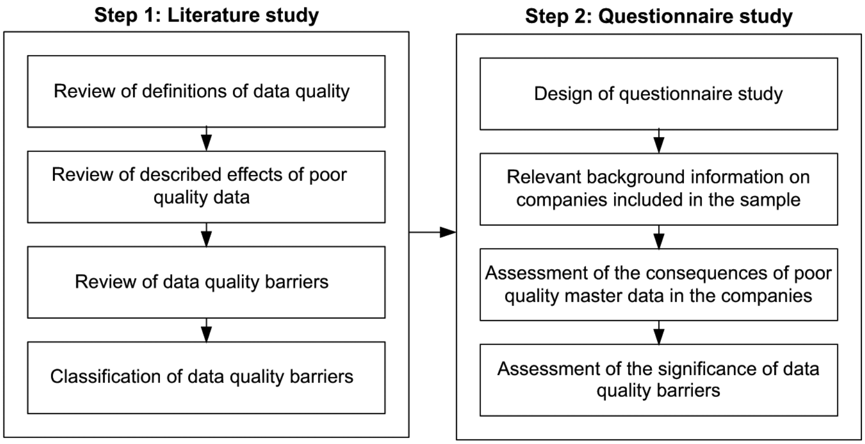

Figure 1.
Research methodology of the paper

2.1 Literature study
To provide a basis for investigating the questions in focus, this section first presents definitions of the term "data quality" and views on the consequences of poor quality data. This is followed by a review of data quality barriers described in literature. In order to provide a clearer picture of relevant data quality barriers, these are classified into a set of overall categories.

2.2 Questionnaire study
To investigate the mutual importance of the defined categories of data quality barriers, a questionnaire study was carried out. In the questionnaire study the focus was limited to master data in order to obtain more exact answers. Thus, the paper only makes conclusions on data quality barriers in relation to master data.

2.2.1 Questionnaire design. The questionnaire consisted of four background questions and 16 questions related to master data. Since the quality of the retrieved data is related to the extent to which the respondent understands what he/she is asked, the questions were phrased plain language and examples that apply to most companies. The questionnaire was created as a webpage for which the contact persons were sent a login name and password. Questions for crosschecking responses were found difficult to include, if not to repeat questions or make the questionnaire too long. Thus, this aspect was not included.

2.2.2 Population and sample. The questionnaire study was carried out in two rounds. The recipients of the first round were the contact persons for their supplier of enterprise resource planning (ERP) system solutions. The recipients of the second round of the study were members of a network of companies with interest in process equipment for the food industry, e.g. component suppliers, food producers, plant suppliers, advisors, etc. These two populations were selected, because of prior collaborations with these, which was hoped to produce higher response rates. Although, the sampling strategy has some resemblance with "convenience sampling", there was not found any reason to believe that these companies are significantly different compared to other industrial companies in Denmark.

2.2.3 Respondents. The questionnaire was aimed at one respondent of each company. In the first round, the questionnaires were distributed to the contact persons in relation to the ERP systems, and in the second round, contact persons from the network of companies related to food industry process equipment. The respondents held job positions such as IT manager, finance manager, and production manager.

2.2.4 Data analysis. The number of companies contacted and the response rates of the study are seen in Table I. As seen in Table I, only a subset of the completed questionnaires was used in this paper. More specifically, it was decided to discard respondents who had answered "I do not know" to all questions relating to consequences of poor quality data and data quality barriers. The number of employees of the companies is shown in Table II. As seen, 46.6 per cent of the companies are small enterprises (0-49) and 69.3 per cent are from 0-199. Thus, SME's (small and medium sized enterprises) constitute a great portion of the sample. The validity of the questionnaire data is subsequently discussed.

The internal validity of a method refers to how accurately it measures what it is supposed to measure. When focussing only on the respondents within the sample in relation to barriers for high master data quality, the retrieved answers are believable. Although, these respondents may have different views on what constitutes a "big data quality barrier", greater barriers would still receive higher rating that less important ones. Also, the two datasets were obtained from two different populations, which made it possible to make crosschecks of answers between the two datasets. These crosschecks showed similar patterns in the answers.

The response rates were 19.2 and 30.6 per cent respectively in the two datasets (see Table I), which is on the low side, although not uncommon (Fowler, 1984).

	Dataset 1		Dataset 2	
	Number of companies	Part of total (%)	Number of companies	Part of total (%)
Questionnaire sent to	858	100.0	180	100.0
Number of received answers	165	19.2	55	30.6
Number of completed questionnaires	64	7.5	41	22.8
Dataset used for this study	54	6.3	36	20.0

Table I.
The questionnaire survey

Number of employees	Dataset 1	Dataset 2	Cumulative total	Cumulative per cent
0-49	18	23	41	46.6
50-199	15	5	61	69.3
200-499	9	3	73	83.0
500-999	4	0	77	87.5
1,000-1,999	5	1	83	94.3
2,000 or more	2	3	88	100.0
Subtotal	53	35		
Do not know	1	1		
Total	54	36		

Table II.
Number of employees

Furthermore, the sample size is somewhat small and therefore has a greater change of sampling error. Also, unit non-response bias (no response from recipient) was not investigated. In this perspective, great uncertainties are associated with the generalizability of the study. On the other hand, if it assumed that, for instance, only 50 per cent of the respondents are capable of answering the questions, the response rate in principle doubles. This may very well be the case, which is illustrated by the low portion of completed questionnaires in the retrieved ones. Analysis of item non-response bias (certain questions not answered) was done by retrieving feedback from some of the respondents who had not completed the questionnaire. This confirmed that many respondents were not able to answer certain questions, which had led them to give up. The high portion of only partly completed questionnaires therefore indicates that only respondents with good insight into their companies answered all the questions. Finally, under the assumption that the relative importance of the individual data quality barriers is similar in most industrial companies, the results clearly indicate that certain barriers are more important than others.

3. Literature review

3.1 Data quality

To discuss the concept of "data quality barriers" an understanding of the concept of "data quality" is necessary. However, many views on this exist. A classic definition of data quality is "fitness for use", i.e. the extent to which some data successfully serves the purposes of the user (e.g. Tayi and Ballou, 1998; Cappiello *et al.*, 2003; Lederman *et al.*, 2003; Watts *et al.*, 2009). Such a definition implies that the concept is contextual or relative. For instance, dimensions of data quality, such as relevance, believability, or usefulness are highly contextual. However, according to Watts *et al.* (2009), models of information quality assessment have tended to ignore the impact of contextual quality on information use and decision outcomes.

In literature, different data quality classifications are found. First, Ballou and Pazer (1985) divide data quality into four dimensions: accuracy, timeliness, completeness, and consistency. They note that the accuracy dimension is the easiest to evaluate, since this is merely the difference between the correct value and what was actually recorded. They argue that the evaluation of timeliness can be carried out in a similar manner. The evaluation of the completeness dimension is also relatively straightforward, as long as the focus is on whether the data is complete or not rather than the degree to which the data is complete. On the other hand, an evaluation of consistency is a bit more complex, since this requires two or more representation schemes for comparison (Ballou and Pazer, 1985). Wand and Wang (1996) limit their discussion to intrinsic data qualities and define four intrinsic dimensions: completeness, unambiguousness, meaningfulness, and correctness. They refer to a comprehensive literature review by Wang *et al.* (1995), which summarizes the most often cited data quality dimensions. This is shown in Table III.

Redman (1998) categorizes the data problems of an enterprise in terms of data view issues (e.g. relevancy, granularity, and level of detail); data value issues (e.g. accuracy, consistency, currency, and completeness); data presentation issues (e.g. the appropriateness of the format, ease of interpretation, etc.); and other issues (such as privacy, security, and ownership). Wang and Strong (1996) divide data quality into four categories:

Data quality dimension	Citations
Accuracy	25
Reliability	22
Timeliness	19
Relevance	16
Completeness	15
Currency	9
Consistency	8
Flexibility	5
Precision	5
Format	4
Interpretability	4
Content	3
Efficiency	3
Importance	3
Sufficiency	3
Usableness	3
Usefulness	3
Clarity	2
Comparability	2
Conciseness	2
Freedom from bias	2
Informativeness	2
Level of detail	2
Quantitativeness	2
Scope	2
Understandability	2

Source: Wang *et al.* (1995)

Table III. Cited data quality dimensions

(1) intrinsic;
(2) contextual;
(3) representational; and
(4) accessibility.

For each category they define a set of 18 dimensions. For the category of intrinsic data, they define dimensions of believability, accuracy, objectivity, and reputation. Haug *et al.* (2009) define three data quality categories: intrinsic, accessibility, and usefulness. Basically, Haug *et al.* (2009) support the definition provided by Wang and Strong (1996) in comparison to others, but they argue that "representational data quality" can be perceived as a form of "accessibility data quality" instead of a category of its own.

Another view on data quality (or at least data properties) is provided by Levitin and Redman (1998). They argue that since processes to produce data have many similarities to processes that produce physical products, data producing processes could be viewed as producing data products for data consumers, a view shared by many others (e.g. Lee and Strong, 2004; Wang, 1998). Thus, Levitin and Redman (1998) discuss how thirteen basic properties of organisational resources translate into properties for data. More database technical perspectives on quality were also found (e.g. Hoxmeier, 1998; Kim *et al.*, 2003).

3.2 Consequences of poor quality data

The consequences of poor data quality in a company may be numerous. First, poor quality data that is not identified and corrected can have significant negative economic and social impacts on an organization (Wang and Strong, 1996; Ballou *et al.*, 2004). The implications of poor quality data are negative effects to business users through lower customer satisfaction, increased operating costs, inefficient decision-making processes, lower performance, and lowered employee job satisfaction (Redman, 1998; Pipino *et al.*, 2002; Kahn *et al.*, 2003). Also, poor data quality increases operational costs, since time and other resources are spent detecting and correcting errors. In companies, data is created and used in all daily operations, data is a critical input in almost all decisions, and data implicitly defines common terms in an enterprise, which makes data a significant contributor to the organization culture (Levitin and Redman, 1998). Thus, poor data quality can have negative effects on the organizational culture (Ryu *et al.*, 2006). Also, poor data quality means that it is impossible to build trust or confidence in the data, and the result may be a lack of user acceptance of any initiatives based on such data (Friedman *et al.*, 2006).

Redman (1998) argues that studies to produce estimates of the total cost of poor data quality have proven difficult to perform. According to Redman (1998), data quality research has not yet advanced to the point in which standard measurement methods are available for any of these issues. On the other hand, Redman (1998) claims that many case studies feature accuracy measures, but he does not provide references nor does he mention if these are academic studies. Redman (1998) also claims that, when measured at the field level, the reported error rates are in the interval of 0.5-30 per cent. Furthermore, according to Redman (1998), at least three proprietary studies have yielded estimates in the 8-12 per cent of revenue range, but informal estimates suggest that 40-60 per cent of the expense of a service organization may be consumed as a result of poor data. Evidence indicates that the economic effect of even small data inaccuracies can be very significant. In this context, Häkkinen and Hilmola (2008) argue that marginal data inaccuracies (e.g. 1-5 per cent) may not necessarily represent a major problem in manufacturing, but that such inaccuracies will have direct effects in terms of lost sales and operational disruptions in the after-sales organizations.

In contrast with the apparent lack of academic journal papers which include multiple company studies of data quality, many industry experts provide such studies. These industry experts include Gartner Group, Price Waterhouse Coopers, and The Data Warehousing Institute, which claim to identify a crisis in data quality management and a reluctance among senior decision makers to do enough about it (Marsh, 2005). Marsh (2005) summarizes the findings from several such surveys, which are presented here:

- 88 per cent of all data integration projects either fail completely or significantly over-run their budgets
- 75 per cent of organisations have identified costs stemming from dirty data
- 33 per cent of organisations have delayed or cancelled new IT systems because of poor data
- \$611bn per year is lost in the US in poorly targeted mailings and staff overheads alone
- According to Gartner, bad data are the number one cause of CRM system failure

- Less than 50 per cent of companies claim to be very confident in the quality of their data
- Business intelligence (BI) projects often fail due to dirty data, so it is imperative that BI-based business decisions are based on clean data
- Only 15 per cent of companies are very confident in the quality of external data supplied to them
- Customer data typically degenerates at 2 per cent per month or 25 per cent annually
- Organisations typically overestimate the quality of their data and underestimate the cost of errors
- Business processes, customer expectations, source systems and compliance rules are constantly changing. Data quality management systems must reflect this
- Vast amounts of time and money are spent on custom coding and traditional methods – usually firefighting to dampen an immediate crisis rather than dealing with the long-term problem.

The results of these surveys clearly illustrate the relevance of better understanding the causes for poor data quality.

3.3 Data quality barriers

Umar *et al.* (1999) presents results from a case study conducted in the telecommunications industry on data quality issues. On this basis, they propose a data quality approach that combines data quality and data architectures into a single framework with a series of steps, procedures, checklists, and tools. In relation to a management perspective on data quality barriers, they mention a set of issues to be considered which can be translated into the possible data quality barriers:

- lack of roles and responsibilities;
- lack of data quality owners;
- inefficient organizational procedures;
- lack of scheduling scenarios;
- lack of reward/reprimand system; and
- neglecting administrative details, e.g. staff training, position descriptions, responsibility shifts, and communication/administrative issues.

English (1999, p. 422) discusses critical success factors for sustainable information quality and reasons why data-cleansing initiatives fail. Among these are:

- lack of training and education;
- lack of incentives; and
- lack of management understanding and active involvement.

Xu *et al.* (2002) summarises the factors influencing data quality which have been described in the literature by 13 authors with publications from 1973 to 1999. This includes:

- training;

- top management support;
- organisation structure;
- change management;
- employee relations; and
- data quality control.

However, the exact meaning or mutual importance of such factors is not discussed.

Lee *et al.* (2006, p. 31) describe an information quality assessment (IQA) that aims to obtain the respondents' subjective assessment of data quality and the respondents' knowledge of the data quality process, programs, and tools that are in place. In section 3 of the IQA, some data quality barriers are listed:

- lack of responsibility for information quality;
- lack of tools;
- lack of appropriate technologies; and
- lack of procedures.

Smith and McKeen (2008) argue that poor management of data leads to "data silos", i.e. companies often manage data at a local level (e.g. department or location), which implies the creation of "information silos" in which data are redundantly stored, managed, and processed. Smith and McKeen (2008) argue that the problem of data silos has been exacerbated in recent years by a number of factors:

- owing to technological capability to store ever-increasing amounts of data, the amount of data has vastly exceeded the organization's ability to manage, analyze, and apply it (Davenport, 2007, p. 154);
- newly acquired enterprise solutions (e.g. ERP, CRM) for data management often unwittingly contribute to further data confusion by adding new layers of complexity to the situation (Fisher, 2007);
- companies often try to solve data problems with half-measured solutions which are ineffective or even counterproductive;
- ownership issues; and
- companies wish to be able to manage data globally, but often end up in workarounds, i.e. short-term solutions which, in the long term, may produce problems.

3.4 Discussion of data quality barriers

The data quality barriers identified can be translated into five overall data quality barrier themes:

(1) lack of delegation of responsibilities for maintenance of data;
(2) lack of rewards for ensuring valid data;
(3) lack of data control routines;
(4) lack of employee competencies; and
(5) lack of user-friendliness of the software used to manage data.

In Table IV, the barriers identified in literature have been organized into the five defined themes. Some of the barriers identified in the literature did not fit directly into the defined themes, but they can be seen as a cause or effect of the defined themes. These barriers are subsequently discussed.

The first of the three identified barriers in literature, which did not directly fit into the defined themes, is "neglecting administrative details". In practice this is translatable to "lack of delegation of responsibilities for maintenance of data" and "lack of data control routines". The next, "change management", may be translated into "lack of delegation of responsibilities for maintenance of data" or "lack of employee competencies", the latter, obviously, in relation to change management. The third, "half measured solutions", can lead to all the themes mentioned and, thus, is also covered by the five themes.

The next section describes the empirical investigation of the defined data quality barrier themes.

4. Questionnaire survey results

In the questionnaire, the companies were asked to evaluate the consequences of poor master data quality in their company. More specifically, the respondents were asked to estimate their level of agreement with the listed consequences of poor master data quality on a scale of 1 (fully disagree) to 5 (fully agree). The answers are shown in Tables V and VI, in which the column furthest to the right shows the average rating on this scale.

As shown, only a small part of the respondents did not find that incorrect master data has any negative effects on their company. In both datasets the use of extra administrative resources was assessed as the main factor among those listed. In fact, the order of importance of the factors is almost identical in the two datasets. The data gives a clear picture of that the consequences of poor quality data are significant. For instance, more than 85 per cent in the two datasets combined agree that poor quality master data to some extent (scores 3-5) implies the use of extra resources at their company, while a bit more than 65 per cent mostly or fully agree on this issue (scores 4-5).

Although "extra use of administrative resources" received the highest scores, the numbers for the other five areas may be even more interesting. All companies experience poor data quality in varying extent. Ideally, such data are corrected before passed on and therefore do not have any negative consequences other that a little administrative work. However, the survey showed that most of the companies in focus experienced consequences relating to extra production recourses, longer lead times, decreased document quality, decreased product quality or decreased performance. Thus, this is a clear indication that data errors are allowed to move through business processes without being corrected until it is too late.

The respondents were asked to assess their agreement with the listed barriers for achieving high master data quality on a scale of 1 (fully disagree) to 5 (fully agree). The answers are shown in Tables VII and VIII.

As seen, "lack of delegation of responsibilities for maintenance of master data" occurs in both datasets as the factor considered most important, while "lack of rewards for ensuring valid master data" was considered least important. In fact, as for the consequences of poor master data, the order of importance is also very similar on the

Table IV.
Data quality barrier themes

Theme	Umar *et al.* (1999)	English (1999)	Xu *et al.* (2002)	Lee *et al.* (2006)	Smith and McKeen (2008)
Lack of delegation of responsibilities for maintenance of data	Lack of roles and responsibilities, lack of data quality owners	Lack of management understanding and active involvement	Lack of top management support, organization structure, employee relations	Lack of responsibility for information quality	Data exceeding the organization's ability to manage them, ownership issues
Lack of rewards for ensuring valid data	Lack of reward/ reprimand system	Lack of incentives			
Lack of data control routines	Inefficient organizational procedures, lack of scheduling scenarios		Data quality control	Lack of procedures	
Lack of employee competencies		Lack of training and education	Training		
Lack of user-friendliness of the software that is used to manage data				Lack of tools, lack of appropriate technologies	Complexity of IT solutions
Cause/effect of some of the above mentioned	Neglecting administrative details		Change management		Half-measured solutions

topic of barriers for master data quality. In fact, around 80 per cent in the two datasets combined agrees that a lack of delegation of responsibilities to some extent (scores 3-5) is a major barrier for ensuring high quality master data, while almost 70 per cent mostly or fully agree on this issue (scores 4-5). The barrier considered second most important is control routines, which around 50 per cent mostly or fully agree on (scores 4-5), while the others barriers were perceived that important in less than 40 per cent of the companies.

The implication of the results is that it seems that a delegation of responsibilities may be the first place to focus if to improve master data quality. The good news is that

(54 respondents) In your company, the consequences of poor master data quality are?	1 (fully disagree)-5 (fully agree) 1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	Do not know (%)	Avg. (1-5)
Extra resources for administration	9.3	0.0	16.7	18.5	51.9	3.7	3.9
Extra production resources	7.4	7.4	16.7	18.5	38.9	11.1	3.4
Longer lead times	9.3	9.3	13.0	22.2	33.3	13.0	3.2
Decreased document quality	7.4	3.7	14.8	27.8	40.7	5.6	3.7
Decreased product quality	11.1	3.7	18.5	22.2	31.5	13.0	3.2
Decreased performance	9.3	3.7	11.1	27.8	42.6	5.6	3.7

Table V. Consequences of poor master data quality – dataset 1

(36 respondents) In your company, the consequences of poor master data quality are?	1 (fully disagree)-5 (fully agree) 1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	Do not know (%)	Avg. (1-5)
Extra resources for administration	2.8	16.7	19.4	19.4	41.7	0.0	3.8
Extra production resources	2.8	13.9	25.0	16.7	25.0	16.7	3.0
Longer lead times	5.6	13.9	25.0	13.9	25.0	16.7	2.9
Decreased document quality	8.3	19.4	25.0	13.9	33.3	0.0	3.4
Decreased product quality	22.2	22.2	16.7	8.3	27.8	2.8	2.9
Decreased performance	8.3	19.4	13.9	19.4	33.3	5.6	3.3

Table VI. Consequences of poor master data quality – dataset 2

(54 respondents) Major barriers to ensuring high master data quality:	1 (fully disagree)-5 (fully agree) 1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	Do not know (%)	Avg. (1-5)
Lack of delegation of responsibilities for maintenance of master data	3.7	7.4	5.6	33.3	38.9	11.1	3.6
Lack of rewards for ensuring valid master data	18.5	11.1	22.2	18.5	13.0	16.7	2.5
Lack of master data control routines	5.6	9.3	20.4	33.3	18.5	13.0	3.1
Lack of employee competencies	7.4	16.7	24.1	18.5	18.5	14.8	2.8
Lack of user-friendliness of the software that is used to manage master data	14.8	18.5	22.2	9.3	24.1	11.1	2.8

Table VII. Barriers for master data quality – dataset 1

(36 respondents) Major barriers to ensuring high master data quality:	1 (fully disagree)-5 (fully agree) 1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	Do not know (%)	Avg. (1-5)
Lack of delegation of responsibilities for maintenance of master data	2.8	11.1	19.4	16.7	47.2	2.8	3.9
Lack of rewards for ensuring valid master data	11.1	36.1	19.4	19.4	8.3	5.6	2.6
Lack of master data control routines	11.1	13.9	25.0	25.0	22.2	2.8	3.3
Lack of employee competencies	16.7	16.7	25.0	22.2	16.7	2.8	3.0
Lack of user-friendliness of the software that is used to manage master data	19.4	22.2	22.2	13.9	19.4	2.8	2.8

Table VIII.
Barriers for master data quality – dataset 2

a delegation of responsibilities for master data quality may be the cheapest of the five barriers to implement (excluding control routines). Even without control routines, the placement of responsibilities for certain data may have a great effect, since when data quality issues eventually are discovered it is clear who did not do their job properly. However, to better ensure that these data quality responsibilities are fulfilled, control routines may be implemented. Extensive control routines may not be necessary, since simply the awareness of that the data quality is controlled may have a motivating effect on the ones responsible. Thus, given the great costs that can be associated with poor quality data, the introduction of simple routines may have high payback rates. The central issue in this context is to identify a proper balance between data quality improvement costs and costs steaming from poor quality data.

When analysing the company size in relation to consequences of poor quality master data or master data quality barriers, no significant correlations were identified.

5. Conclusion

There seems to be a general consensus that poor quality data is a costly problem for many companies. This paper aims at identifying the main barriers to ensuring high data quality and clarifying which of these are the most important. Since the ways in which master data and transaction data are created and used are fundamentally different, the empirical part of the paper focuses only on master data in order to produce more precise results. Based on a literature review, the paper defines five overall data quality barriers:

(1) Lack of delegation of responsibilities for maintenance of master data.

(2) Lack of rewards for ensuring valid master data.

(3) Lack of master data control routines.

(4) Lack of employee competencies.

(5) Lack of user-friendliness of the software that are used to manage master data.

To investigate the relevance of the defined data quality barriers from a master data perspective, the paper presents a parts of a questionnaire study of two rounds. The first round of the questionnaire study resulted in 54 useful answers and the second round in 36. Both datasets showed that almost all the companies believe that they experience significant negative effects when encountering poor quality master data. In both datasets, the use of extra administrative resources was given the highest rating, followed by decreased document quality and decreased performance. The interesting aspect about this part of the survey is that if poor quality of master data is corrected instantly, this only results in additional administrative costs. However, since the companies reported other experienced consequences of poor quality master data, it seems that poor quality data only to a small extent are checked and corrected before it is too late.

Next, the respondents were asked to consider the importance of the five defined master data quality barriers encountered during operations. The answers in both datasets showed that the barrier, which is believed to have the greatest effect, is "lack of delegation of responsibilities for maintenance of master data". This is followed by "lack of control routines", while "lack of rewards for ensuring valid master data" was considered least important. The implication of the results is that delegation of responsibilities may be the first place to focus if the objective is to improve master data quality. The paper argues that, besides being a relative cheap initiative (if not including control procedures), the placement of responsibilities for certain data alone may have a great effect, since when data quality issues eventually are discovered someone is responsible. However, initiating other data quality initiatives may further help raise data quality levels. The central issue when choosing initiatives is to identify a proper balance between data quality improvement costs and costs steaming from poor quality data.

For both the consequences of poor master data and barriers for master data quality, the order of importance is similar in the two datasets. This supports the validity of the findings. On the other hand, the way the sample was selected and the low response rates raises some questions regarding the generalizability of the sample. However, although the data retrieved from the questionnaire are not strong enough to make generalising claims, they provide strong indications of certain barriers being more important than others.

The contributions of this paper, i.e. identification of barriers for ensuring high master data quality and providing empirical support of the importance of these, represent a step towards a better understanding of how to raise the level of master data quality in companies. Thus, the paper provides a better understanding of the barriers for master data quality, which can help companies in the creation of master data quality policies and other data management related tasks. Furthermore, the paper may stimulate future research on the topic, including better solutions for dealing with data quality problems. Given the significant negative consequences of poor quality data, which many companies experience, such work may have a great impact on the efficiency and competitiveness for such companies.

References

Ballou, D. and Pazer, H. (1985), "Modeling data and process quality in multi-input multi-output information systems", *Management Science*, Vol. 31 No. 2, pp. 150-62.

Ballou, D.P., Madnick, S. and Wang, R. (2004), "Assuring information quality", *Journal of Management Information Systems*, Vol. 20 No. 3, pp. 9-11.

Cappiello, C., Francalanci, C. and Pernici, B. (2003), "Time-related factors of data quality in multi-channel information systems", *Journal of Management Information Systems*, Vol. 20 No. 3, pp. 71-91.

Davenport, T. (2007), *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, Boston, MA.

English, L.P. (1999), *Improving Data Warehouse and Business Information Quality*, John Wiley & Sons, New York, NY.

Fisher, T. (2007), "Demystifying master data management", *CIO Magazine*, 30 April.

Fowler, F.J. Jr (1984), *Survey Research Methods*, Sage, Thousand Oaks, CA.

Friedman, T., Feinberg, D., Beyer, M.A., Gassman, B., Bitterer, A., Newman, D., Radcliffe, J., White, A., Paquet, R., DiCenzo, C., Logan, D., Blechar, M., Knox, R.E., Bell, T. and Shegda, K.M. (2006), *Hype Cycle for Data Management, 2006*, GartnerGroup Research, July 6, ID#G00140057.

Häkkinen, L. and Hilmola, O-P. (2008), "ERP evaluation during the shakedown phase: lessons from an after-sales division", *Information Systems Journal*, Vol. 18 No. 1, pp. 73-100.

Haug, A., Pedersen, A. and Arlbjørn, J.S. (2009), "A classification model of ERP system data quality", *Industrial Management & Data Systems*, Vol. 109 No. 8, pp. 1053-68.

Hoxmeier, J.A. (1998), "Typology of database quality factors", *Software Quality Journal*, Vol. 7 Nos 3-4, pp. 179-93.

Kahn, B., Strong, D. and Wang, R. (2003), "Information quality benchmarks: product and service performance", *Communications of the ACM*, Vol. 45 No. 4, pp. 184-92.

Kim, W., Choi, B-J., Hong, E-K., Kim, S-K. and Lee, D. (2003), "A taxonomy of dirty data", *Data Mining and Knowledge Discovery*, Vol. 7 No. 1, pp. 81-99.

Knolmayer, G. and Röthlin, M. (2006), "Quality of material master data and its effect on the usefulness of distributed ERP systems", *Lecture Notes in Computer Science*, Vol. 4231, pp. 362-71.

Lederman, R., Shanks, G. and Gibbs, M.R. (2003), "Meeting privacy obligations: the implications for information systems development", *Proceedings of the 11th European Conference on Information Systems (ECIS), Naples, Italy, 16-21 June*, available at: http://is2.lse.ac.uk/asp/aspecis/20030081.pdf (accessed 29 June 2009).

Lee, Y. and Strong, D. (2004), "Knowing-why about data processes and data quality", *Journal of Management Information Systems*, Vol. 20 No. 3, pp. 13-39.

Lee, Y.W., Pipino, L.L., Funk, J.D. and Wang, R.Y. (2006), *Journey to Data Quality*, MIT Press, Cambridge.

Levitin, A.V. and Redman, T. (1998), "Data as a resource: properties, implications, and prescriptions", *Sloan Management Review*, Vol. 40 No. 1, pp. 89-101.

Madnick, S., Wang, R. and Xian, X. (2004), "The design and implementation of a corporate householding knowledge processor to improve data quality", *Journal of Management Information Systems*, Vol. 20 No. 1, pp. 41-9.

Marsh, R. (2005), "Drowning in dirty data? It's time to sink or swim: a four-stage methodology for total data quality management", *Database Marketing & Customer Strategy Management*, Vol. 12 No. 2, pp. 105-12.

Pipino, L.L., Lee, Y.W. and Wang, R.Y. (2002), "Data quality assessment", *Communications of the ACM*, Vol. 45 No. 4, pp. 211-18.

Redman, T.C. (1998), "The impact of poor data quality on the typical enterprise", *Communications of the ACM*, Vol. 41 No. 2, pp. 79-82.

Ryu, K.-S., Park, J.-S. and Park, J.-H. (2006), "A data quality management maturity model", *ETRI Journal*, Vol. 28 No. 2, pp. 191-204.

Smith, H.A. and McKeen, J.D. (2008), "Master data management: salvation or snake oil?", *Communications of the Association for Information Systems*, Vol. 23 No. 4, pp. 63-72.

Tayi, G.K. and Ballou, D.P. (1998), "Examining data quality", *Communications of the ACM*, Vol. 41 No. 2, pp. 54-7.

Umar, A., Karabatis, G., Ness, L. and Horowitz, B. (1999), "Enterprise data quality: a pragmatic approach", *Information Systems Frontiers*, Vol. 1 No. 3, pp. 279-301.

Vayghan, J.A., Garfinkle, S.M., Walenta, C., Healy, D.C. and Valentin, Z. (2007), "The internal information transformation of IBM", *IBM Systems Journal*, Vol. 46 No. 4, pp. 669-84.

Wand, Y. and Wang, R.Y. (1996), "Anchoring data quality dimensions in ontological foundations", *Communications of the ACM*, Vol. 39 No. 11, pp. 86-95.

Wang, R.Y. (1998), "A product perspective on total data quality management", *Communications of ACM*, Vol. 41 No. 2, pp. 58-65.

Wang, R.Y. and Strong, D.M. (1996), "Beyond accuracy: what data quality means to data consumers", *Journal of Management Information Systems*, Vol. 12 No. 4, pp. 5-34.

Wang, R.Y., Storey, V.C. and Firth, C.P. (1995), "A framework for analysis of data quality research", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7 No. 4, pp. 623-40.

Watts, S., Shankaranarayanan, G. and Even, A. (2009), "Data quality assessment in context: a cognitive perspective", *Decision Support Systems*, Vol. 48 No. 1, pp. 202-11.

Xu, H., Nord, J.H., Brown, N. and Nord, G.D. (2002), "Data quality issues in implementing an ERP", *Industrial Management & Data Systems*, Vol. 102 No. 1, pp. 47-58.

About the authors

Anders Haug, PhD is an Associate Professor in the Department of Entrepreneurship and Relationship Management, University of Southern Denmark. His research focuses on knowledge engineering, information systems and knowledge management. He has published a long list of papers on these topics in international journals, such as *Industrial Management & Data Systems*, *International Journal of Physical Distribution & Logistics Management*, *International Journal of Advanced Manufacturing Technology*, *International Journal of Industrial Engineering*, *Journal of Intelligent Manufacturing* and *Management Research News*. He has practical experience from projects in relation to software design, process optimization, knowledge/information management and product design.

Jan Stentoft Arlbjørn is a Professor of Supply Chain Management in the Department of Entrepreneurship and Relationship Management at the University of Southern Denmark in Kolding. He holds a PhD in Logistics and Supply Chain Management. His research and teaching interests are within supply chain management, relationship management, enterprise resource planning, strategic management, corporate social responsibility and supply chain innovation. He has published research articles in journals such as *International Journal of Physical Distribution & Logistics Management*, *Supply Chain Management: An International Journal*, *Journal of Operations Management*, *Operations Management Research*, *Industrial Management & Data Systems*, *International Journal of Production Economics*, *International Journal of Integrated Supply Management*, *Journal of Agile Systems and Management*, *European Business Review*, and *Industrial and Commercial Training*. He primarily instructs MSc, PhD and MBA students in supply chain management, operations management and administrative information systems. He has practical experience from positions as Director of Program Management Office at LEGO System A/S, Axapta ERP Project Manager at Gumlink A/S and management consulting. Jan Stentoft Arlbjørn is the corresponding author and can be contacted at: jar@sam.sdu.dk