Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2011 Proceedings

European Conference on Information Systems (ECIS)

Summer 10-6-2011

CAN DATA QUALITY HELP OVERCOME THE PENGUIN EFFECT? THE CASE OF ITEM MASTER DATA POOLS

Maria Madlberger

CAN DATA QUALITY HELP OVERCOME THE PENGUIN EFFECT? THE CASE OF ITEM MASTER DATA POOLS

Madlberger, Maria, Webster University Vienna, Berchtoldgasse 1, A-1220 Vienna, Austria, madlberger@webster.ac.at.

Abstract

The diffusion of standards is characterized by network effects, path dependency, and the penguin effect. Particularly the latter, also referred to as excess inertia, is a frequent inhibitor of the adoption of standards, even if they could provide benefits. This is particularly true for item master data pools that suffer from little adoption in many industries as benefits can only accrue if many firms use them. At the same time, data pools show the potential to improve the quality of item master data by pooling the efforts on data quality assurance. This paper addresses the question whether an improvement of item master data quality can contribute to overcoming the penguin effect by data pools. The theoretical considerations are supplemented by an exploratory qualitative research among the leading retailers in the Austrian food and drug sector. The findings suggest that data quality improvement can be one way to encourage the use of data pools and thus overcome the penguin effect in adoption.

Keywords: Item Master Data, Data Quality, Data Pools, Global Data Synchronization Network, Penguin Effect, Excess Inertia, Standardization

1 Introduction

The effectiveness of interorganizational information sharing strongly depends on the quality of the shared data (Hartono et al., 2010). This is particularly true for sharing of item master data, i.e., data on product attributes, such as identification data, size, weight, and price. Item master data is necessary for logistics processes in supply chains as they control the physical flow of goods, inventory management, transportation, and the space management in the retail outlets. It is also needed for the financial flow, e.g., invoice control and payment. The quality of master data directly impacts the efficiency of transactions as errors in these data have a detrimental effect on all related transactions. In supply chains the points of origin and use of item master data differ: the manufacturers of goods are the originators of item master data, but the retailers who distribute the goods are in need of them. Therefore item master data has to be shared between manufacturers and retailers.

For several years the concept of item master data pools has been discussed as a means to centralize sharing of item master data. Instead of multiple bilateral exchanges of data a pool acts like a clearing center: manufacturers transfer the data to the pool only once and retailers collect the data from the pool. The centralization of the data flow allows the centralization of quality control, too. In contrast, a bilateral exchange of data requires individual quality control at each manufacturer-retailer dyad. Thus, by replacing multiple individual data quality control efforts, data pools can allow a stronger focus on data quality control at one single point. This should help improve the quality of item master data.

However, in practice data pools are used only to a little extent and the quality of item master data is unsatisfactory. A recent report shows that 80% of an industry's item master data is inconsistent which results in GBP 47m annual costs of corrections and manual workarounds for British grocery retailers and suppliers. These figures are supplemented by another projected GBP 95m shrinkage costs and GBP 60m lost sales per year caused by poor data quality. Therefore the estimated damage to the British grocery industry due to insufficient data quality is as high as GBP 202m per year (GS1 UK, 2009). The use of data pools is growing, but still sporadic in many countries and industries, especially among retailers. For example, among more than 23,000 data pool users identified by the Global Location Number, only 360 are retailers (Garry, 2010). One main reason is a classical startup problem in the adoption of standards, the so-called penguin effect or excess inertia (Farrell and Saloner, 1985; Farrell and Saloner, 1986). As data pools are largely subject to network effects (Katz and Shapiro, 1994), companies hesitate to pioneer in adoption to avoid the risk of making the investment, but not being able to accrue benefits if other firms do not adopt the pool. This stand-off locks firms into inefficient bilateral item master data exchange. Often firms even prefer manual data exchange by fax or spreadsheets attached to e-mails. As known for more than 20 years, manual data handling is extremely error-prone (Dearing, 1990). The high involvement of manual work can reduce the quality of item master data substantially and therefore is a cause of the above-mentioned low level of data quality.

The paper at hand is motivated by the lack of master data pool adoption in practice. Given the importance of high quality of master data and the clearly stated data quality improvement potentials of data pools, the paper particularly seeks to reveal whether data quality can be a driver that is strong enough to overcome the penguin effect as an adoption barrier to data pool usage. If this is the case, large inefficiencies can be avoided if a data pool secures a higher level of data quality than in the present situation. As a further consequence, the attractiveness of a data pool as an enabler of data quality improvement could become larger than the hindering penguin effect and the related initial investments. Besides the research stream of standards adoption, the paper further investigates the role of perceived data quality as an antecedent of information systems use (Wright and Donaldson, 2002, Fletcher and Wright, 1995, Payton and Zahay, 2003).

Currently little empirical research on master data management and its drivers is available. To empirically address this question, the paper discusses the results of a qualitative study among leading retailers and wholesalers in a particular industry, i.e., the Austrian food and drug sector. This industry undertook an ambitious effort to adopt item master data pools almost ten years ago. After a short-term “hype” the implementation largely failed and left retailers and manufacturers disappointed and with substantial sunk costs. Since then, the exchange of item master data is taking place manually across the entire industry which is diametrically opposed to the large penetration of EDI-based transaction data exchange. The exploratory study sheds light on the role of data quality in an environment where the penguin effect is extremely strong due to a high frustration level and large sunk costs. The findings reveal that the role of data quality for master data pool adoption is more complex than assumed. Instead of being a “simple” driver, data quality has an ambiguous impact on data pool acceptance and adoption intention. The study further shows that data quality of master data consists not only of “absolute” dimensions such as accuracy or up-to-dateness, but also relative issues like the congruence between the manufacturers’ and the retailers’ data attributes and data structures.

The paper is organized as follows: section two outlines the basics of item master data exchange. Section three discusses theory-based adoption barriers of data pools. After the presentation of the research methodology, the results of the exploratory study are presented. Four key questions are addressed: the benefits and costs of data pools, the need for data quality improvement, the perceived contribution of data pools to quality improvement, and the ability of data quality improvement to overcome the penguin effect. The conclusion outlines future research directions in this under-researched area.

2 Interorganizational item master data exchange

2.1 The importance of high item master data quality

As supply chain processes are associated with the flow of goods, data that identifies and specifies product items are needed. Such data has the character of master data and is referred to as product information (Legner and Schemm, 2008), item or product data (Nakatani et al., 2006), or core product data (Popa and Duica, 2010). To account for the particular characteristics of master data and differentiate it from transaction data on the one hand and distinguish it from other kinds of master data (e.g., location or customer master data) the data in question is termed “item master data” in this paper.

The quality of item master data is of utmost importance, especially if managed in one single database source (Smith and McKeen, 2008). Item master data is used in almost all business activities and departments. Each transaction that involves a firm’s products is affected by item master data. If errors occur, they affect all transactions that follow the entry of the wrong data parameters. The severity of errors in item master data can be illustrated by the following scenario that can cause a retailer large costs: A retailer receives item master data with wrong size information, i.e., the data indicates a smaller size than the physical product has. If the wrong data is related to the transportation unit (e.g., size of carton), the error will be detected at the receipt of the goods in the warehouse, as the carton does not fit into the designated space. In this case, a larger space must be found and the data must be corrected once. If, however, the wrong information is related to the consumer unit, wrong space management decisions may result and the products may not fit into the shelves of the store outlet. In the worst case, a large retailer with several hundreds of stores will have to find a larger shelf space in each store. Furthermore the retailer must correct the wrong data and transfer the correction to all affected stores. Thus, for a retailer, errors in master data can quickly be multiplied by the number of store outlets, thus also multiplying the time and costs of error correction.

Data quality is measured along various dimensions. Seddon (1997) applies relevance, timeliness, and accuracy of information generated by an information system to operationalize data quality. Rai et al. (2002) measure information quality along the dimensions of precision of information, the output that is exactly needed, sufficient information to complete the task, absence of errors, accuracy, and helpfulness for the related problem(s). In their update of the original DeLone and McNeal model

(DeLone and McNeal, 1992), DeLone and McNeal (2003) report that studies investigating the construct of information quality measured it along the criteria accuracy, up-to-dateness, completeness, relevance, and consistency. Up-to-dateness in considered important as item master data must be available when the physical products are distributed. Completeness is relevant as retailers have to request for missing information and complete it on their own. Relevance refers to the fact that retailers do not need huge amounts of master data attributes that go beyond their business requirements, but just what they need for their business processes. Finally, consistency is a particularly critical issue. As a recent UK-based empirical study shows, the consistency of item master data attributes in the grocery industry is alarmingly low. An analysis of almost 18,000 unique items showed that the dimension mis-match between two retailers was 82% and that between three retailers 98%. In other words, data of three retailers on one attribute of the same item were identical in 2% of the cases (GS1 UK, 2009).

2.2 Global Data Synchronization and data pools

In a traditional setting, item master data is stored separately at each organization in the supply chain where they “maintain their own versions of data about the items they handle” (Nakatani et al., 2006, p. 7). From the interdependencies of resources viewpoint, however, item master data are a typical example of a pooled interdependency (Thompson, 1967; Kumar and van Dissel, 1996). Such interdependency is characterized by resources that are shared between organizations with high mutual independence elsewhere and a high degree of structurability. Thus the necessary interaction between the involved organizations can be minimal (Robey and Sales, 1994).

To overcome the inefficiencies caused by redundancy and multiple work, industry initiatives set up an infrastructure for synchronizing item master data between organizations. A substantial contribution was made by GS1, a global non-profit organization that originates from the merger of the Uniform Code Council (UCC) and EAN (European Article Numbering) Europe. The data synchronization infrastructure is named Global Data Synchronization (GDS) and is intended to overcome the shortcomings of bilateral item master data exchange particularly among many trading partners. In consistence with the above-mentioned pooled interdependencies, GDS considered master data pools (in brief data pools) that serve as electronic intermediaries between manufacturers and retailers. Examples of data pools are the b2b exchanges 1SYNC and Agentrics (Legner and Schemm, 2008). The German market is served by the data pool SINFOS that merged with the Agentrics pool GenSync in 2007 and thus became the global data pool SA2 Worldsync (SA2 Worldync, 2010). Like any electronic intermediary, data pools are subject to strong network effects (Legner and Schemm, 2008), which requires a high critical mass of involved items and participants to become attractive. Therefore instead of competing single data pools an interoperability of the existing data pools was headed for which resulted in the Global Data Synchronization Network (GDSN). GDSN consists of two main components: the data pools and the global registry that links the data pools.

2.3 Contribution of data pools to an increase in item master data quality

By centralizing the exchange of item master data, data pools can contribute to an increased data quality. This can be effected in two ways: (1) by reducing the manual work of exchanging non-integrated item master data that needs to be re-keyed by the retailers’ purchase departments and (2) by actively providing data verification and quality control services.

Data pools can reduce manual work by providing structured and integrated electronic item master data instead of paper-, spreadsheet- or pdf-based master data sheets. Many data pools handle the transfer of item master data in the EDIFACT standard PRICAT (price and sales catalog) that can be directly imported into a materials management system (MMS) or enterprise resource planning system (ERPS) (Legner and Schemm, 2008). Thus, the error-prone procedure of re-keying data (Dearing, 1990) is restricted to individual additions made to the data by the retailer.

Various studies empirically demonstrate the savings potentials of GDS, mediated by improved data quality. A study by Accenture conducted among various U.S.-based brand manufacturers reveals considerable savings potentials through GDS, e.g., up to 50% higher improved productivity within the order and item administration for retailers and 23% decrease of time from item entry to the retailer's shelf (1SYNC, 2006). A case study on Wal-Mart shows that after having implemented GDSN, item maintenance was decreased from 15-30 days down to one day. Wal-Mart could further reduce out-of-stocks by 2.5% (GS1 Australia, 2010). Furthermore, as manual work can be cut, the saved personnel capacities can be applied for increased quality control which can further improve the quality of item master data.

The second way of data pools' contribution to increase data quality is the offering of data validation services (Schemm and Legner, 2008). As the pool ideally is a central hub that connects many manufacturers with many retailers, a data validation service can be achieved at the pool itself as one single source of data for a whole supply chain. Instead of having many retailers doing more or less the same work of data quality control, the data pool can perform this task once and thus only distribute high quality data to retailers. Data quality control consists of automated validation rules and plausibility checks, but also physical verification by measurement and item inspection (Schemm and Legner, 2008). An extensive data quality check is offered by the Swedish data pool Validoo item that is operated by GS1 Sweden. This data pool runs a lab where 30 key item master data attributes are checked (GS1 Sweden, 2010). As up-to-dateness of item master data is a key component of data quality, data pools can also contribute by ensuring a fast distribution of data to the recipients.

3 Adoption barriers to data pool usage

The diffusion of standards follows several rules identified in literature. As data pools require a standardized exchange of item master data, these rules apply for their adoption, too. The most important factors that influence standard adoption are network effects, standardization costs, penguin effects, and path dependency (Zhu et al., 2006; Weitzel et al., 2006). For item master data pools, direct and indirect network effects (Katz and Shapiro, 1994) are relevant. To be attractive for retailers, the pool needs to be used by many manufacturers to share a large portion of data through it. To be attractive for manufacturers, many retailers must use the pool to justify the effort. Indirect network effects are associated with additional services, particularly by improvement of data quality. Standardization costs are characterized by showing an asymmetry between the quantification of costs versus benefits. While the costs of standardization can easily be estimated prior to standard adoption, benefits are often hard to assess and will only be measurable some time after implementation (Weitzel et al., 2006). This also occurs with data pool usage: the license and adaptation costs are easily measurable, but benefits due to improved item data quality are difficult to quantify.

The term penguin effect was introduced by Farrell and Saloner (1986) and describes the phenomenon of excess inertia. This effect refers to a risk-avoiding behavior by single firms when it comes to pioneering a standard adoption. There is no incentive for one individual firm to adopt a standard on its own as it is exposed to the risk of investing into the standard without getting any benefit in return if others do not adopt it. Farrell and Saloner (1986) drew an analogy with animal behavior: penguins do not want to be the first to enter the water for finding food therefore they wait until others go first. In doing so, they minimize the risk of being hunted by predators. While, however, penguins sooner or later will have to enter the water because they will be forced to do so by starvation, firms that consider the adoption of item master data pools may not feel an increasing pressure to adoption. To overcome this excess inertia, there need to be some strong incentive that induces firms to take action. Existing poor data quality could be a motivation to adoption that is larger than the desire for risk avoidance.

Perceived poor data quality itself is an adoption barrier to information systems use. Wright and Donaldson (2002), Fletcher and Wright (1995) and Payton and Zahay (2003) investigated poor information quality as a barrier to systems adoption. In an e-government context, Gilbert et al. (2004) confirmed that information quality is a significant system adoption barrier. As an antecedent of IS success information quality was investigated in the structural models provided by DeLone and McNeal (1992), Seddon (1997), and the updated model by DeLone and McNeal (2003). The proposed models were further investigated by Rai et al. (2002) in the context of a quasi-voluntary setting and by Wang (2008) in the context of e-commerce systems. Venkatesh and Davis (2000) incorporated output quality into the technology acceptance model and revealed that this factor impacts intention to use mediated by perceived usefulness.

As master data is of core importance to retailers, they undertake large efforts to control the quality on their own. In the past, data pools largely provided data of low quality as many suppliers did not care sufficiently for quality control of data uploaded to the data pool. The result was an "electronic exchange of bad data" (Vuyyuru et al., 2005, p. 2). Due to several factors, for example synchronization issues, the difficulties of agreeing on a single definition of every data item, changing legal and regulatory considerations, and security issues, the achievement of a high data quality is a major challenge to a data pool (Smith and McKeen, 2008). Thus firms' distrust in the capabilities of a centralized data quality control hinders their acceptance of a data pool.

4 Research methodology

To shed light on this underresearched field, an exploratory research design with in-depth interviews (Yin, 1994) is chosen in this research. To address the issue of interorganizational master data exchange, the study is focused on item master data that is continuously provided by manufacturers to retailers. In this initial stage of the research on master data quality the focus lies on the perspective of retailers and wholesalers as the institutions that receive and use master data. The study is conducted among the leading retailing firms in the groceries and drug supply chain of Austria. The country and industry were chosen because of the almost hundred percent degree of manual, bilateral master data exchange that totally lacks centralized master data quality control. Furthermore, previous efforts of an industry-wide introduction of a German item master data pool failed in the past due to a low number of participants and unsatisfactory performance. The resulting stand-off resembles exactly the penguin effect described in literature.

Prior to the interviews, the author developed an extensive interview guideline comprising more than 30 questions on current practices and problems of master data exchange, requirements on an electronic exchange of master data, and attitude toward a master data pool. The guideline and selection of the retailers were discussed with two executives and one specialist of GS1 Austria as well as the editor of a leading trade journal who is a food and drug industry expert. Eight leading food and drug retail and wholesale companies were selected for the study. Among them are the market leaders and the second strongest firms in the Austrian retail and wholesale business. Prior to the interviews the executives of the chosen eight firms were asked for participating in the interviews and coordinating further separate or joint interviews with experts from the purchasing, IT, and supply chain management departments. All contacted retail companies agreed to participate in the survey.

To collect the data in an exploratory way, personal in-depth interviews were conducted. The data collection comprises eleven in-depth interviews (nine personal interviews, one telephone interview, and one telephone conference). The interview partners are 18 executives and heads of IT, purchasing, and supply chain management departments. Prior to the interviews, further nine executives and department managers who could not join the interviews provided input to the respondents. The total duration of the interviews was 26 hours. Thus on average one interview was conducted with 1.63 persons and took more than two and a half hours. Eight of the eleven interviews were tape recorded for further analysis, in the remaining three interviews the respondents declined recording. After transcription of the interviews, parts of the answers were reorganized in case they better fitted to a different question. After this step, the data was consolidated, i.e., all firm's answers were put together and sorted by the questions. This step allowed identifying commonalities and differences between the interviewees' positions.

5 Results and discussion

This section presents the findings of the qualitative study in respect of the role of data quality for the adoption of item master data pools by retailers. The results are organized along four dimensions: (1) the benefits and costs of data pools as stated by the retailers, (2) the present pressure for item master data quality improvement, (3) the perception whether item master data pools can improve item master data quality improvement, and (4) the question whether item master data quality is a factor that can help overcome the penguin effect.

5.1 The benefits and costs of data pools

While literature clearly states the benefits that are associated with data pools, the interviewed retailers propose a more differentiated picture of the benefits, but also the costs of data pools. In contrast to the argumentation in literature, retailers refer to much more specific and process-related positive and negative impacts of data pools. Table 1 shows the benefits and costs of data pools that were mentioned by the majority of the interviewed retailers.

Benefits of data pools	**Relevant to number of firms (max. 8)**
Support of a complex master data management	7
Higher data quality	6
Facilitation of international business relationships	6
Lower logistics and marketing costs	6
More up-to-dateness of data	5
Facilitation of internal master data management	5
Costs of data pools	**Relevant to number of firms (max. 8)**
High IT investments	7
Requires internal adaptations	7
Is only profitable if adopted by many trading partners	6
Insufficient discipline by suppliers	6
Poor distribution by data pool operator in Austria	5

Table 1: Benefits and costs of data pools considered relevant by retailers

5.2 The pressure for item master data quality improvement

High item master data quality is an absolute must for all interviewed retailers. This issue has such a high priority that all retailers perform manual data quality control after receipt of the data from the manufacturers. Data quality control is laborious as only few errors can be detected by plausibility checks. In particular, there are different categories of data errors that differ in their degree of severity. Errors which are too small for being identified by plausibility checks can turn out to be dangerous. As one interview partner pointed out:

> If the volume is 100 milliliters but the data say 100 liters, the error is obvious and will not cause any further trouble. If data indicate 120 milliliters, the error is much more difficult to detect. However this can result in severe trouble. From a legal viewpoint the error can be interpreted as fraud and this may lead to lawsuit, large image loss and damage to the firm.

On the other hand, some errors will automatically be detected when products proceed in the supply chain. Differences in measurement turn out to problems in the inventory and invoice differences are detected at invoice verification and payment at the latest. Nevertheless, as the severity of errors cannot be predicted, firms put much effort on quality control.

A key source of poor data quality is the manual entry of item master data. However, with one exception, firms state that the frequency and dimension of these errors is within an acceptable range. A larger issue, however, is the data quality provided by many manufacturers, especially small and medium-sized ones. These firms often do not have the necessary information technology capabilities to provide complete data. Often it is also caused by a lack of discipline as several retailers state. A typical case is described by the following statement:

> Master data is often provided in an incomplete manner. It depends on the suppliers: large firms are o.k., but small suppliers or importers from Far East are often problematic. They do not have the necessary IT. Examples of missing data are dimensions, country of origin, or best before date. Sometimes suppliers provide only part of the information, for example only the remaining shelf-life, but not the best before data, as requested by us. To "educate" our suppliers we installed a supplier evaluation system that is also an input for our annual appraisals.

Data quality is also related to the handling of identification numbers. If suppliers do not provide a separate EAN number for each modified item, retailers need to find ways to distinguish an old version of an item from a new one:

> Multinationals assign a new EAN number to each new or modified item. This makes the item continuous in our processes. Smaller firms, however, sometimes use the same EAN for several product variants. In such a case we can decide whether we treat the variants as if they were identical or we need to rework the data.

To improve suppliers' discipline, retailers exert pressure on suppliers by conditions and multiple requests. These measures, however, are dependent on the individual retailers' power. Larger retailers can impose more power, but cannot achieve receipt of fully complete data from all suppliers either.

Despite the mentioned problems with data quality, retailers do not feel a strong pressure for data quality improvement. They perceive to have a high quality level of data after their own manual quality control although this is very laborious. Quality control covers re-measurement of physical dimensions (sometimes done in the warehouse), comparison of keyed data with original data file, comparison with similar items or copying fields from related items, or even photography of items to supplement missing product images. The costs of these control processes could hardly be estimated by the retailers and wholesalers. They are more or less considered an inevitable necessary process that cannot be replaced. Austrian food and drug retailers and wholesalers have assortments ranging from 8,000 to 50,000 product items with up to another 30,000 short-term seasonal articles per year. Given this dimension the total workload of quality control is significant.

5.3 The contribution of data pools to item master data quality improvement

The firms were explicitly asked whether they perceive a contribution of data pools to an improvement of data quality. The answer to this issue is twofold. On the one hand, based on their prior experience most interviewed firms argue that existing data pool systems did not provide a satisfactory quality of item master data so far. One the other hand, among eight firms, seven are convinced that data pools can improve data quality. Only one firm, the largest doubter on data pools in general, states that data pools cannot improve data quality at all.

The experience of poor data quality provided by data pools is closely associated with a significant related issue. This major issue is the heterogeneity of data content and structure among retailers. At present each retailer and wholesaler requests item master data by their individual item master data sheets that vary considerably across the industry. The mismatch between data content and structure provided by a data pool and the individual requirements is characterized by the following statement:

> We did not use a data pool because the data quality is not sufficient, data is not complete, our internal data structures do not fit to external ones, and an automated integration into the ERP systems is not useful. […] With bilateral data exchange, we receive the data the way we need it. […] It is not really poor quality, but the data pool does not deliver the data in the structure we need. […] Of course each

> supplier wishes to transfer item master data in a standardized way around the world once, but if retailers need different units, e.g., measuring units or weights, suppliers have to account for this fact.

As the retailer states this problem is not related to per se wrong data, however the heterogeneity of individual needed data structures prevents an exchange of data in a way it is needed. Even if data provided by the pool is correct, complete, and controlled, it does not have a satisfactory quality for an individual retailer or wholesaler.

A solution to this problem requires either a standardization of required data structures and content or a large number of different data fields provided by the data pool. According to the interviews, the second approach is more realistic as retailers and wholesalers are unlikely to change their internal data requirements in order to meet external data standards. On the other hand, a data pool can offer many alternative data structures and allow individual customization of the user interface.

Except the above-cited retailer that is very doubtful about data pools, all other retailers and wholesalers strongly agree that data pools can improve data quality. The main contribution is seen in the centralization of quality control that reduces multiple efforts and the standardization of the quality control procedure. As one retailer executive points out:

> Transparency of data and a uniform quality level are important. For retailers a quality control performed by a data pool is attractive as the pool checks the data of different suppliers in the same way. Thus the pool enables a standardization of quality control as is ensures that each retailer gets data with the same, high quality. This quality assurance is a clear added value.

Several retailers and wholesalers further state that less workload on in-house quality control frees capacities that can be used for more data quality improvement in challenging product categories. For example, fruits and vegetables are rather complex in terms of their master data, thus increased efforts in this area could lead to noticeable process improvements.

All interviewed firms clearly state that the current decentralized data quality control will not be totally replaced by a centralized quality control, at least not in the short run. They require the possibility to view and control the data before they are integrated into the internal information system. Furthermore, as each retailer and wholesaler has to supplement the item master data by internal attributes (e.g., individual texts, storage locations, or responsible persons) an interface for manual data entry before transfer to the internal system is necessary anyway.

5.4 The role of data quality to overcome the penguin effect

The situation in the observed industry contains a paradox: on the one hand, data pools are expected to improve data quality, on the other hand, the penguin effect prevents any adoption. High data quality is extremely important to retailers and wholesalers. This fact calls for and hinders adoption of data pools simultaneously. While the interviewed firms see an opportunity to make quality more effective by data pool usage, they hesitate to outsource data quality control at the same time. Nevertheless, as most retailers consider external data quality a value added, even if it is supplemented by additional in-house quality control, there is some possibility that data quality improvement can at least weaken the penguin effect.

One wholesaler executive even argues that a sound quality control is the key to the adoption of data pools by retailers and wholesalers. He recommends GS1 and data pool operators to invest large efforts and personnel resources to achieve a sound quality control procedure that is proven by extensive testing.

> It is a chicken-egg-problem and GS1 is the chicken who must stand up for the pool. Together with the pool operator, GS1 needs to provide a solution that ensures a sound data quality. This can only be done with significant personnel resources and marketing efforts. It is clear that this step implies an entrepreneurial risk. However, only then it will be possible to achieve a larger commitment among retailers that is necessary for a broad adoption in the industry.

Following the analogy of Farrell and Saloner (1986), this firm expects the data pool operator and GS1 to be the first penguins entering the water. If they ensure a high data quality, other retailers will commit themselves to data pool usage and the penguin effect could be overcome.

For the majority of the firms, however, data quality is one challenge for data pool adoption besides various others. Table 2 presents the challenges data pools need to overcome for adoption.

Challenging factor	Exemplary statements made in interviews
High implementation costs of the data pool	"Requires much energy to adapt internal interfaces." "Costs and time efforts are substantial."
Selection of appropriate suppliers and product categories: trade-off between appropriateness of standardized cate-gories and low improvement potentials	„Large, stable suppliers of categories such as nutriments and categories with less innovations […] are more appropriate [for a data pool], however in these categories the suppliers are also proper in the existing system."
Exchange of price data is too sensitive for an external data pool	„It is hard to administer price changes via a data pool as this issue is too sensitive. But many data changes are based on price changes."
Achievement of an industry-wide commitment of large retailers, wholesalers, and manufacturers	"A commitment of all retailers and manufacturers is needed. It must contain a binding time schedule, otherwise the whole issue peters out again"
Internal adaptations are required	"The data pool turns the manufacturers' obligation to provide data into the retailers' obligation to collect data. This increases retailers' responsibility"

Table 2. Challenges for data pool adoption besides data quality

While all retailers agree that data pool implementation cannot be initiated alone even by the largest retailer, the majority of the surveyed managers argues that a joint commitment among the large retailers, wholesalers, and manufacturers creates the prerequisites for a large-scale implementation. Based on the aggregate suggestions of the interviews, an implementation scenario should consist of the following steps:

(1) Select an appropriate data pool.

(2) Achieve the commitment of several large retailers. Possible approaches are a joint decision with or without a binding time schedule, a large pilot project, a large test, and the installation of a task force. Affected actors, such as purchasing departments, should be involved.

(3) Achieve the commitment of large manufacturers of appropriate product categories.

(4) Quick implementation and near-term performance measurement.

In this scenario, data quality becomes relevant at several stages. In the first stage, data quality can be formulated as a key requirement on the data pool system. In the second and third stage, data quality improvement can be an incentive to undertake joint efforts and establish an industry-wide commitment for the adoption of a data pool. Finally, in the fourth stage the performance measurement can clearly reveal the impact on master data quality.

6 Conclusion

The exploratory study presented in this paper indicates that the interrelations between data quality and the adoption of a data pool are complex and reciprocal. On the one hand, negative experiences with data quality provided by a data pool in the past resulted in the penguin effect. On the other hand, most retailers and wholesalers anticipate an improvement of data quality through the use of a data pool. However, the perceived risk of pioneering the data pool adoption alone and thus not gaining any benefits because of network effects is currently larger than the expectations on benefits through improved item master data quality.

Nevertheless, individual large retailers have started initiatives to GDSN adoption. Ahold undertook large efforts to encourage its suppliers to transfer their item master data to a data pool (Gallagher, 2005). The U.S. retailer Wegmans and the supplier Wakefern strongly support the use of data pools to increase the accuracy and therefore the quality of item master data (Garry, 2010). The U.S. foodservice business recently launched an initiative for adopting business communication standards including the use of data pools (Holzman, 2010). The attractiveness of the Swedish data pool Validoo is largely driven by its data quality checking service (GS1 Sweden, 2010). Thus for individual firms, but also some industries, data quality is a strong argument for the usage of data pools.

This research reveals the relevance of data quality as a factor of the adoption of data pools as a jointly applied interorganizational system with large network effects. To get a deeper understanding of the issue of data pool adoption, additional research is needed. First, as data pools serve retailers and manufacturers, the standpoint of the latter need to be revealed. Although retailing firms often can exert their power, there must be an understanding under which conditions manufacturers are willing and able to adopt a data pool. A second point is the investigation of possible further factors that may drive or inhibit data pool adoption. Emerging consumer trends, such as the request for more detailed product information, but also commercial innovations based on technology development (e.g. extended product information available via mobile Internet) will increase the number and dimensions of item master data. A third issue is representativeness of the empirical research that requires a quantitative survey as a supplement to the qualitative exploratory research. This is not only true for the geographic scope of research, but also for investigated industries.

Acknowledgment: The author thanks GS1 Austria for its comprehensive support of the empirical study.

References

1SYNC (2006). Synchronization – The next generation of business partnering. Available online at http://www.1sync.org/documents/news/Industry%20Studies/SynchronizationNextGenofBizPartnering.pdf, accessed 2010-06-11.

Dearing, B. (1990). The strategic benefits of EDI. Journal of Business Strategy, 11 (1), 4-6.

DeLone, W.H. and McLean, E.R. (1992). Information systems success: The quest for the dependent variable. Information Systems Research, 3 (1), 60-95.

DeLone, W.H. and McLean, E.R. (2003). The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems, 19 (4), 9-30.

Farrell, F. and Saloner, G. (1985). Standardization, compatibility, and innovation. Rand Journal of Economics, 16 (1), 70-83.

Farrell, F. and Saloner, G. (1986). Installed base and compatibility: innovation, product preannouncements, and predation. The American Economic Review, 76 (5), 940-955.

Fletcher K. and Wright G. (1995). Organizational, strategic and technical barriers to successful implementation of database marketing. International Journal of Information Management, 15 (2), 115-126.

Gallagher, S. (2005). Ahold starts data sync trials via divisions. Supermarket News, 53 (9), 51-55.

Garry, M. (2010). Syncing up. Supermarket News, 58 (26), 62-64.

Gilbert, D., Balestrini, P., and Littleboy D. (2004). Barriers and benefits in the adoption of e-government. International Journal of Public Sector Management, 17 (4), 286-301.

GS1 UK (2009). Data crunch report. The impact of bad data on profits and consumer service in the UK grocery industry. Available online: http://www.gs1uk.org/what-we-do/sector-solutions/retail/Pages/Master-Data-Management.aspx, accessed 2010-06-11.

GS1 Australia (2010). GDSN retailer benefits. Available online: http://www.gs1au.org/assets/documents/products/gs1_system/gs1_gpc_gdsn_retailer_benefits.pdf, accessed 2010-06-11.

GS1 Sweden (2010). How Validoo Q-lab works. Available online: http://www.gs1.se/en/Validoo/Validoo-Q-lab/How-Validoo-Q-lab-works/, accessed 2010-06-12.
Hartono, E., Li, X., Na, K.-S., and Simpson, J.T. (2010). The role of the quality of shared information in interorganizational systems use. International Journal of Information Management, 39 (5), 399-407.
Holzman, J. (2010). Global standards improve supply chain. Food Logistics, (125), 44-46,
Katz, M. and Shapiro, C. (1994). Systems competition and network effects. Journal of Economic Perspectives, 8 (2), 93-115.
Kumar, K. and van Dissel, H.G. (1996). Sustainable Collaboration: Managing Conflict and Cooperation in Interorganizational Systems. MIS Quarterly, 20 (3), 279-300.
Legner, C. and Schemm, J. (2008). Toward the inter-organizational product information supply chain – evidence from the retail and consumer goods industries. Journal of the AIS, 9 (3/4), 119-150.
Nakatani, K., Chuang., T.-T., and Zhou, D. (2006). Data Synchronization technology: Standards, business values and implications. Communications of the AIS, 17, 2-60.
Payton, F.C. and Zahay, D. (2003). Understanding why marketing does not use the corporate data warehouse for CRM applications. Journal of Database Marketing, 10 (4), 315-326.
Popa. V. and M. Duica (2010). Global standards and best practices for supply chain information alignment in consumer goods and retail. In Proceedings of the 4th European Conference on Information Management & Evaluation (de Castro Neto, M. Ed.), p. 335, Universidade Nova de Lisboa, Lisbon.
Rai, A., Lang, S.S., and Welker, R.B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. Information Systems Research, 13 (1), 50-69.
Robey, D. and C.A. Sales (1994). Designing Organizations. 4th Edition. Richard Irwin, Homewood, IL.
SA2 Worldsync (2010). About SA2 Worldsync. Available online: http://gds.sa2worldsync.com/gds/about_SA2/index.html.en, accessed 2010-05-28.
Schemm, J. and C. Legner (2008). The role and emerging landscape of data pools in the retail and consumer goods industries. In Proceedings of the 41st Hawaii International Conference on System Sciences (Sprague R.H. Ed.), p. 150, IEEE Computer Society, Los Alamitos, CA.
Seddon, Peter B. (1997). A respecification and extension of the DeLone and McLean model of IS success. Information Systems Research, 8 (3), 240-253.
Smith, H. and McKeen, J.D. (2008). Developments in practice XXX: Master data management: Salvation or snake oil?. Communications of the AIS, 23 (4), 63-72.
Thompson, J.D. (1967). Organizations in Action: Social Science Bases of Administration. McGraw-Hill, New York, NY.
Venkatesh, V. and Davis, F.D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science, 46 (2), 186-204.
Vuyyuru, P., Upadhye, A., and Faldu, T. (2005). Global data synchronization alone is not enough. White paper, available online: http://www.infosys.com/offerings/industries/retail/white-papers/Documents/global-data-synchronization.pdf, accessed 2010-06-11.
Wang, Y.-S. (2008). Assessing e-commerce systems success: a respecification and validation of the DeLone and McLean model of IS success. Information Systems Journal, 18 (5), 529-557.
Weitzel, T., Beimborn, D., and König, W. (2006). A unified economic model of standard diffusion: The impact of standardization cost, network effects, and network topology. MIS Quarterly, 30 (Special Issue), 489-514.
Wright, G. and Donaldson, B. (2002). Sales information systems in the UK financial services industry: An analysis of sophistication of use and perceived barriers to adoption. International Journal of Information Management, 22 (6), 405-419.
Yin, R.K. (1994). Case Study Research: Design and Methods. 2nd Edition. Sage, London.
Zhu, K., Kraemer, K.L., and Gurbaxani, V. (2006). Migration to open-standard interorganizational systems: Network effects, switching costs, and path dependency. MIS Quarterly 30 (Special Issue), 515-539.