
Data quality assessment and improvement

Risto Silvola, Janne Harkonen*, Olli Vilppola,
Hanna Kropsu-Vehkaperä, and Harri
Haapasalo

Industrial Engineering and Management,
University of Oulu, Finland, P.O. Box 4610,
FI-90014 University of Oulu, Finland
e-mail: risto.silvola@pp2.inet.fi
e-mail: janne.harkonen@oulu.fi
e-mail: ollivilppola@gmail.com
e-mail: hanna.kropsu-vehkaperä@oulu.fi
e-mail: harri.haapasalo@oulu.fi
*Corresponding author

Abstract: Data quality has significance to companies, but is an issue that can be challenging to approach and operationalise. This study focuses on data quality from the perspective of operationalisation by analysing the practices of a company that is a world leader in its business. A model is proposed for managing data quality to enable evaluation and operationalisation. The results indicate that data quality is best ensured when organisation specific aspects are taken into account. The model acknowledges the needs of different data domains, particularly those that have master data characteristics. The proposed model can provide a starting point for operationalising data quality assessment and improvement. The consequent appreciation of data quality improves data maintenance processes, IT solutions, data quality and relevant expertise, all of which form the basis for handling the origins of products.

Keywords: data quality; assessment; improvement, operationalisation; master data; product data management; information systems; dimensions; operations; business.

Biographical notes: M.Sc. Risto Silvola received his Masters degree in Mechanical Engineering from University of Oulu, Finland. He has worked for leading global manufacturers at several different positions both in Business organisation and in IT. During his career, he has participated in new product programs, global business change programs in program management positions. He is working with his long-term dream, postgraduate studies, where he tries to combine 20 years of experience working on the field with theory. His aim is to build up a solid big picture of data management of products/services.

Dr Janne Harkonen received his Bachelor's degree (1st Class Honours) in Engineering Business Management from University of Greenwich in the UK and both his M.Sc. in Process Engineering and Dr (Tech) in Industrial Engineering and Management, from the University of Oulu, Finland. He has also studied in the University of North Carolina at Wilmington, USA. He has worked for many years in the IT and environmental technology industries, in the UK and in Finland. Currently, he is a senior research fellow at the University of Oulu. Dr Harkonen has authored and co-authored over thirty journal articles and also a number of other publications.

Mr Olli Vilppola received his M.Sc (Tech) in Industrial Engineering and Management in 2014 from the University of Oulu. Currently he is working at the case company as a data quality analyst. His main responsibilities include Master Data Management from the data quality perspective, such as data quality profiling, ETL-development and data quality monitoring.

Dr Hanna Kropsu-Vehkaperä received her M.Sc. in 2005, and her Dr (Tech) in Industrial Engineering and Management in 2012, both from the University of Oulu, Finland. Currently she is a post-doctoral research fellow and the Head of Industrial Engineering and Management undergraduate and graduate programmes at the University of Oulu, Finland. Her research interests include product development, product data management, and especially the management of product development projects through product structure, and requirements.

Dr Harri Haapasalo is the Professor of Product Management at Industrial Engineering and Management, University of Oulu in Finland. He has two main areas of research; one in product management, product development and second one in area of complex systems management, lean construction and business models. Professor Haapasalo has been very active in obtaining research projects, and active in journal publications. His list of publications contains more than 200 international items. He has supervised more than twenty doctoral theses and been external examiner for more than twenty doctoral dissertations.

1 Introduction

Technological developments have resulted in companies handling increasing amounts of data. Even though the significance of data quality is understood to an extent, considerable unnecessary costs and challenges are faced by many companies (Haug *et al.* 2011; Smith, 2011). Data quality involves many aspects that make operationalising the concept very challenging (Alpar and Winkelsträter, 2014). Data quality consists of different characteristics, or dimensions that have significance for effectively addressing relevant matters (e.g. Tayi and Ballou, 1998).

Poor data quality has an impact on many areas of business and operations (e.g. Puurunen *et al.*, 2014; Hazen *et al.*, 2014; Glowalla and Sunyaev, 2014). One of the challenges of data quality involves change, due to the data in databases requiring synchronisation with the surrounding world to remain fit for purpose (Orr, 1998). Overall, poor data quality can have severe impacts on the entire enterprise, resulting in increased costs, customer dissatisfaction, and even influencing employee morale (Haug *et al.*, 2011; Silvola *et al.*, 2011; Redman, 1998). Hence, data quality is not just an IT problem, but is closely related to organisational issues (Haug *et al.*, 2013).

Calculating the costs caused by inadequate data quality is challenging, due to the indirect nature of these costs (Eppler and Helfert, 2004). Consequently, a different approach may be required. For example, Loshin (2011) argues that determining the impacts of data quality on business processes can be manageable through analysing smaller components and lost value. Adequately understanding and managing product data is seen to have clear business benefits (Otto, 2012a).

Focusing on operating with valid master data can help in improving performance (Arlbjørn *et al.*, 2007). Better focus on key business objects at different levels and stages can help practitioners to improve data quality (Ofner *et al.*, 2013). Some relevant case studies exist that concentrate on master data management and organisational aspects (e.g. Otto, 2012b). In fact, data quality is seen as a success factor for business processes (Ofner *et al.*, 2012).

Different frameworks have been developed for data quality in general, the one by Wang, and Strong (1996) probably being the most cited, focusing on the needs of data consumers. Eppler and Wittig (2000) presented a review of data quality frameworks and concluded that a data quality framework ought to have four capabilities including evaluation, analyse, solving issues, and form basis for measurement and proactive management. Many authors have discussed different frameworks, whilst the development and the development trends of data quality management are well documented (Xiao, *et al.*, 2014). However, previous frameworks have been somewhat challenging from the perspective of truly operationalising data quality. Effective assessment and improvement of data is a necessity for companies, taking the realities of complex information systems and relevant processes (Madnick *et al.*, 2009). Actions towards data quality must meet the business needs and user expectations to form a solid basis for measuring data quality (Gustafsson *et al.*, 2006).

This study analyses the state of data management and data quality from the perspective of operationalising data quality. This is done by both, analysing the existing literature and relevant practices in a company that is a world leader in its business. In addition, a model is proposed for managing data quality to enable evaluating and operationalising data quality.

The above described can be condensed into the following research questions:

RQ 1 What are the data quality challenges and requirements for data?

RQ 2 How to ensure data quality for product data management?**2 Literature review**

Many researchers view data quality as fitness for use (Ofner *et al.*, 2012; Breur, 2009; Tayi and Ballou, 1998; Wang and Strong, 1996). Hence, data quality is dependent on the actual use of the data (Wang and Wang, 1996). There is a tendency that the same data has multiple uses and users (Shankaranarayanan and Cai, 2006; Tayi and Ballou, 1998). This can potentially lead to a situation where the data quality is experienced high by some, and low by others (Sundararaman, 2011). Hence, data quality can have dual characteristics and be subjective, i.e. meet expectations and objective, i.e. meet specifications (Eppler, 2006). Consequently, data quality is context dependent and a multidimensional concept (Ofner *et al.*, 2012).

The importance of understanding different dimensions of data quality and their significance for effectively addressing data quality issues has been emphasised (e.g. Tayi and Ballou, 1998). Wang (1998) introduced a product perspective on data quality management and recognised that also an information product has many perspectives. The difference to a physical product is that data is used by multiple users and does not deplete while being used. There can be strong positive interrelationships among data quality dimensions, influencing corporate capabilities (Kwon *et al.*, 2014). Hence, the concept of dimensions can provide a frame of reference for data quality measurement by providing different perspectives for data quality considerations (Loshin, 2011). However, only those dimensions that prove feasible to the surrounding industrial setting can truly be utilised by practitioners (Batini *et al.*, 2007).

Several authors have introduced and discussed dimensions for data quality. These dimensions can have quality categories such as accuracy, timeliness, completeness, consistency, change frequency, and many others. (Ballou and Pazer, 1985; Wang and Strong, 1996; Van Zeist and Hendriks, 1996; Leung, 2001; Kahn *et al.*, 2002; O'Donoghue *et al.*, 2011; Hüner *et al.*, 2011; Rocha, 2012; Alpar and Winkelsträter, 2014). For example, Kahn *et al.* (2002) present sixteen dimensions for data quality, including: Accessibility, Appropriate amount, Believability, Completeness, Concise representation, Consistent representation, Ease of manipulation, Free-of-error, Interpretability, Objectivity, Relevancy, Reputation, Security, Timeliness, Understandability, and Value added. Nevertheless, the four most common ones presented in the literature seem to be Accuracy, Consistency Completeness and Timeliness (e.g. Weber *et al.*, 2009; Blake and Mangiameli, 2011).

Poor data management may result in companies managing data locally, causing redundant storage, management, and processing of data (Smith and McKeen, 2008). Also traceability can be lost as a result of poor data management (Miksa *et al.*, 2014). Marginal data quality issues can also have an impact, for example affecting after-sales organisations more while functions such as manufacturing can live with them (Häkkinen and Hilmola, 2008). Hence, also smaller data issues may have significance. Company characteristics also have an influence on how issues are experienced (e.g. Motiwalla and Li, 2013). In general, the lack of delegation of responsibilities for maintaining master data, and lack of routines are identified amongst the aspects that have the greatest impact on master data quality (Haug and Arlbjørn, 2011).

Data quality management has become increasingly important, both from the academic and professional perspectives due to the increasing awareness of the impacts of data

quality. Data quality is seen to have even higher priority than for example usability and services interaction (Shahin *et al.*, 2014). This quality-oriented approach understands data as an asset and values it through different stages from provisioning to disposal (Weber *et al.*, 2009). Common characteristics can be found in most notable data quality management approaches. For example, Batini *et al.* (2009) names the common three to include analysing the state, assessment/measurement, and improvement.

Wang (1998) was one of the first, and is among the most cited frameworks for data quality management, derived from the well-known total quality management philosophy and representing a product perspective on data quality. The approach supports organisations in institutionalising continuous improvement of information product (Batini *et al.*, 2009). Also other frameworks have been presented, including total information quality management, a framework by English (2003), designed for data warehouse projects. Virtuous cycle of data quality is another example, one that aims to a situation where data quality practitioners provide feedback that leads to data quality improvements (Loshin, 2011). Also fuzzy analytical hierarchy process is used to evaluate data quality dimensions (Khosroanjom *et al.*, 2011). As Batini *et al.* (2009) stated, many similarities can be found in different data quality management approaches, and they all can be divided into at least two main areas, namely assessment and improvement. The execution of different frameworks varies to some extent, some are more ongoing than others, and business impacts are acknowledged differently. Some frameworks also pay more attention on justification and creating commitment for data quality efforts (e.g. Gustaffson *et al.*, 2009; Eppler and Helfert, 2004). For example, Wang (1998) developed a survey-based diagnostic instrument for data quality assessment and indicated how total data quality management can be applied. In addition, multi-aspect data quality assessment has been discussed (Kulikowski, 2014). When companies are choosing a suitable data quality assessment method, focusing strictly on a particular approach may not result in meeting all the needs, but an alternative approach may be required (Woodall *et al.*, 2013).

Even though data quality management requires processes, tools and people to function, it is often people's perspective that is neglected (Wende, 2007). More particularly, roles and responsibilities are handled in a general manner, even though they play a critical role in data quality management (Otto *et al.* 2007). The efforts are often solely driven by IT, ignoring organisational issues, whilst master data management implementation problems are rather organisational and process based than technology based (Silvola *et al.*, 2011). Khatri and Brown (2010) refer to data governance as "what decisions must be made to ensure effective management and use of IT, and who makes the decisions". This is to focus on management and use of IT to achieve corporate performance goals. Also Weill and Ross (2004) emphasise how data governance should not be considered in IT isolation, due to its link to other organisational assets. In addition, Weber *et al.* (2009) point the importance of business perspective to data governance as an enabler for the implementation of organisation-wide responsibilities for data quality management, including both IT and business departments. Data governance can assist in defining the relevant roles and responsibilities. The responsibility assignment matrix provides a potential frame for the roles (Wende, 2007). Also Loshin (2011) defines similar roles with focus more on strategically aligned organisation-wide data quality. In fact, the RACI (responsible, accountable, consulted, informed) matrix is proposed by many authors (e.g. Loshin, 2011; Weber *et al.*, 2009; Ferreira *et al.*, 2014; Wende, 2007). Product data management can have a role in implementing policies, procedures and guidelines defined through data governance.

When planning data quality improvements, current state is important, but one should not forget that organisations capability to improve also plays a factor (Caballero *et al.*,

2008). Obtaining understanding on the level of maturity that best meets the needs can provide a criterion to analyse an organisation's capability (Loshin, 2011). Separating different data quality goals can help organisations to achieve their data quality objectives through management (Caballero *et al.*, 2004). The capability aspects are discussed in the literature to some extent (e.g. Paulk *et al.*, 1993; Hüner *et al.*, 2009; Ryu *et al.*, 2006; Chuang and Lin, 2014).

3 Research process and case company description

Figure 1 illustrates the research process. Relevant literature on product data management and data quality were first analysed to reveal the state of managing data quality and the currently utilised methodologies. Analysing the literature provided an overview on how data quality is understood, how organisations address the relevant issues, and how different authors have contributed. Data quality components of assessment and improvement were both considered. The case company was selected based on the opportunity to have an access to a company that is a world leader in its business. The company operates in business-to-business markets providing equipment and services to a variety of industries. The company has more than ten thousand employees, operations in a number of countries and revenue of several billions. Data and data quality are very relevant considerations for the case company due to the nature of their product portfolio and service offering. Data relevant to the case company's products and services is tracked and controlled by IT processes involving; Product Data Management (PDM), Enterprise Resource Planning (ERP), Customer Relations Management (CRM), Global Service offering, company internal employment and organisation data, and a sales configuration platform. The company also has some relevant ongoing initiatives and harmonisation efforts. The case company uses one of the world's most widely used product lifecycle management (PLM) software for managing product master data. The case company analysis was further focused on those data domains that can be seen to have master data characteristics and involve multiple processes at the case company. The analysis involved different perspectives. Data quality issues were analysed from the perspectives of item data, company data, people data, service/asset data and supply chain management. Interview questionnaire was created based on the understanding obtained through the literature analysis.

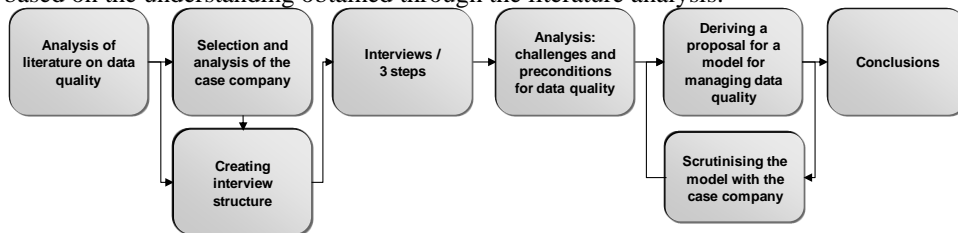


Figure 1. Research process

The study included eighteen interviews involving various process owners, product managers, heads of IT, assistant controller, IT architects, Master data managers and PDM specialists. The interviews were threefold to identify any data quality issues and assess their business impact; assess the relative importance of different type of data quality issues; and to assess the maturity of data quality management and the case company. The interview steps: 1) Interviewees were asked to describe the most common data quality issues in their domain, including potential root causes and business impacts, 2) the identified data quality

issues were rated against 16 data quality dimensions (Kahn *et al.*, 2002) by the interviewees. The rating was done on a 10-point Likert scale, with 1 the lowest, 5 average, and 10 the highest importance. This was done to obtain requirements for data in different data domains. The sixteen data quality dimensions were further classified according to PSP/IQ (Lee *et al.*, 2002) model to illustrate how different data domains would like to see their data, 3) Clarifying current data quality capabilities by reviewing the case company's state of data quality management. The intentional variety enabled obtaining a broad view over data quality with coverage from management level to user interface. Finally, a model was proposed for managing data quality to enable evaluating and operationalising data quality.

The research utilised a semi-structured thematic interview approach (Merton *et al.*, 1990), whilst interviews were conducted informally, in a qualitative manner to allow the interviewees to explain and clarify the cases and topics as entities. All the interviews were recorded to enable thorough analyses, whilst the interviewer made complementary notes during the interview sessions. Based on analysing the interview material and the literature, a model was derived for managing data quality. The model was scrutinised together with the case company to ensure the quality of the proposed model. Finally, conclusions were made together while considering the implications of this study.

4 Results

Data quality issues were analysed from the perspectives of *item data*, *company data*, *people data*, *service/asset data* and *supply chain management*.

Item data consists of data relating to case company's products. Each item has a specific several-digit case company ID (CC-ID), a unique identifier that links all item attributes. Examples of item attributes include; version, specification, description, main group/procurement category, producer, item group/classification, item type, magnitude/unit of measure, item owner, status, criticality and basic technical attributes. PDM system is used to manage the case company's item data. For enterprise resource planning purposes, the PDM system is connected to company ERP system.

Company data consist of data on customers, vendors and the case company. The data is managed through a centralised repository. The data in the centralised repository consists of basic information that identifies the company, including name, address, VAT number, phone, Data Universal Numbering System number, legal hierarchy information, cross-reference information for different systems and the company type. The centralised repository distributes needed basic company information into company systems, such as ERP and CRM.

People data are personnel master data that consists of employee information, including name, address, bank account, employment information and organisation information. The case company's financial department owns the people data that is located in a specific system.

Asset and Service data are closely related and relevant to field service processes. Asset data includes data on customer equipment assemblies, including information on who has provided the equipment. Hence, the asset data describes the current and available data on the current installed equipment base. Service data, on the other hand, includes all data within the entire service delivery process. Service data covers new equipment installations and maintenance according to service requests.

Supply Chain management perspective to data quality is deeply related to overall master data quality. Along the entire supply chain, different data are added according to

customer orders, including Item data, Bill of Materials, customer information, delivery times and so forth. Hence, various data should be integrated properly to have a seamless supply chain.

4.1 Data quality challenges and requirements

The data quality related challenges and requirements were *discussed in more detail for one data domain, namely Item data*. The data quality challenges and derived requirements for all the analysed data domains are presented in the sub-chapter “synthesis”.

Item data

The analysis of data quality dimensions for item data revealed how the interviewees emphasised different viewpoints of item data. It was seen of highest importance that item data is securely maintained, easily accessible, originates from a credible source, is accurate and free-of-errors, whilst being beneficial and provides advantages from its use. Nevertheless, data quality dimensions of completeness, interpretability, reputation, and timeliness were rated nearly as important in relation to item data. Item/ product data are seen to consist of sensitive data that provide the case company advantages in the markets, and hence are confidential due to the importance for the business.

The interviewees placed particular importance on data security especially when there is presence in emerging markets. This is described by an interviewee comment: *“it is number one priority to secure the main assets from the competitors”*. Noteworthy is that accessibility can be seen as a contradiction to security, both of which were rated highly important for item data. The importance of both dimensions can be explained by an interviewee comment, *“data has to be globally visible for its lifecycle point of view, such as the product spare parts management”*. The importance of access to item data and product configurations/ product structure by the means of Bill of Material (BOM) was emphasised critical particularly from the manufacturing perspective. The believability, the extent to which data is credible, correct and trusted by others, raises from multiple dimensions. According to the interviewees *“certain attributes should be 100 percent accurate, but with some attributes the accuracy is not so important”*. However, in the context of ERP, *“difficulties with the accuracy can mean 3-4 times more work than with the correct data”*. Hence, more important attributes are seen to require more focus. Timeliness is seen to have a similar logic. The amount of item data is not viewed as critical in terms of appropriate volume for the information. On the other hand, the data manipulation ought to be minimal in the case of master data.

The Item data quality dimensions that were found important are presented within PSP/IQ model (Figure 2). The PSP/IQ model indicates that the requirements for the desired item data quality should emphasise dependability and usability aspects.

		Conforms to Specifications	Meets or Exceeds Consumer Expectations
Product Quality	<u>Sound information</u>		<u>Useful Information</u>
	Free-of-error/accuracy	7.04	Appropriate amount 4.55
	Concise representation	4.14	Relevancy 6.21
	Completeness	6.42	Understandability 6.21
	Consistent Representation	5.80	Interpretability 6.42
	5.85	Objectivity 4.97	5.67
Service Quality	<u>Dependable Information</u>		<u>Usable Information</u>
	Timeliness	6.42	Believability 8.07
	Security	7.87	Accessibility 7.04
			Ease of Manipulation 5.18
			Reputation 6.42
	7.15	Value-Added 7.25	6.79

Figure 2. Item data quality classification PSP/IQ

Data quality challenges relevant to Item data can be seen as context-independent, data perspective issues. *Duplicates* are seen as the biggest Item data quality challenge. In the case company, duplicates are caused by previous use of multiple ERP systems and legacy IDs. Nevertheless, the case company is continuously removing duplicates and enforcing CC-IDs and the past burden is in the process of being removed. Duplicates have versatile impacts in many areas of the case company's business. At worst, the consequences of duplicate Item data can be critical. For example, duplicates cause difficulties for calculating the total volume of purchased items, influencing reporting. Also, duplicates can result in missed business opportunities if items are purchased at wrong prices, or if wrong spare parts are purchased. An interviewee described the situation with duplicates as: "*the company cannot achieve global savings by making better contracts, because the annual volume cannot be calculated when multiple IDs exist for a single item*". The less severe impacts of duplicates include the related disorder, additional data maintenance needs, and extra work being required for finding the correct data.

Missing data is another major Item data related challenge. For example, ERP implementations at different sites may result in missing data. This can be due to an item being opened in a wrong place. Missing data can cause negative impacts on productivity, slow down processes, or stop production. The interviewees had a clear estimate of the level of financial impact of missing data. The existence of *irrelevant data* was also seen as a challenge when databases contain inactive items that would need to be removed. Irrelevant data involves the product life cycle management considerations. An interviewee described the situation with irrelevant data as: "*there is no technique to remove the data that is not in use, and select the relevant and active data when doing the reporting*".

The case company has ongoing efforts to improve their data quality. Item data belongs to a data domain that is aware of the importance of *managing data quality*. For example, it is quite clear for them that ERP system requires certain level of data quality to provide full benefits. The case company is well aware of the above discussed data quality challenges and have taken action to address them. One of the actions involves introducing a data quality tool that enable checking technical data in company systems, including indicators for inconsistencies, duplicates accuracy, timelines and completeness. The company is also utilising a defined service level agreement for the indicators. Nevertheless, detecting data flaws is only a part of the work, as *validating Item data* still requires manual efforts. Aside the data quality tool, data quality policies and data standards are currently more up-to-date.

The case company has global processes for creating Item data to ensure that Items are created in accordance with strict rules and data dictionaries along specific standards. All data attributes are clearly defined together with presenting valid values and relevant examples. The case company also has rules for maintaining, changing and correcting Item data. Training is also provided globally for different stakeholders. Nevertheless, there are still some perspectives of managing Item data quality that require further attention.

The *reaction times* on data quality issues are seen to be still too long, yet at tolerable levels. Currently, users and relevant business units are notified by the following day of identifying data issues. The notification includes instructions for correcting the data. Root causes are typically analysed, especially in the case of newly created Item data, to enable improving future data quality. Unfortunately, tight schedules and lack of resources sometimes result in some data issues going undetected until the production. The case company has recognised conducting *business impact analysis* on Item data as an important issue, even if it is not actively carried out. Nevertheless, data profiling on Item data is done on ad-hoc basis. One of the biggest challenges of managing the quality of Item data in the case company is seen to involve *data governance*. The case company has gone through significant effort to assign executive level responsibilities, initiated a global team for data quality, and nominated relevant data owners. Governance as a term however is experienced to have negative connotation as it is linked to activities that no one is truly responsible. The case company has learned that governance is more effective and better accepted once named differently. For example, the case company utilises *data cluster steering group meeting* for similar activities. In practice, the global business owner chairs this meeting, and the head of IT has an advisory role. In this way both IT and business have a clear role and responsibilities are clearly visible. Consequently, long-term success of the efforts requires stability in terms of personnel and increased *resources*. This is to avoid unnecessary fire-fighting mode and enable more systematic data quality work and to strengthen continuous improvement.

Synthesis

Table 1 presents the data quality challenges, and derived data quality requirements that were found the most important for each data domain in the case company. It can be seen that the challenges and derived requirements have similarities, but also some differences. The most common *data challenges* across the data domains involve *duplicates, irrelevant, missing, or outdated data*. The data domain of supply chain management is however, somewhat different. In supply chain management, Item and Company data play a critical role by providing the “content” and the “address” for the orders. Consequences of *design errors* become visible in the form of wrong parts in assembly, and additional purchases caused by change requests. *Master data issues* are caused by a dilemma of changes not being ideal anymore in the data domain of supply chain management, but change requests forcing to make changes to different data.

Table 1. Data quality challenges and derived requirements

Data Domain	Data Challenges	Data Quality Management Challenges	Derived Data Quality Requirements
Item data	Duplicates	Validating Item data	Believability
	Missing data	Managing irrelevant data	Security
	Irrelevant data	Reaction times	Value-Added
		Business impact analysis	Accessibility

		Data governance Resources	Accuracy
Company data	Duplicates Irrelevant data Missing reference data	Measuring data quality	Believability
		Reporting	Relevance
		SLAs	Reputation,
		Data standards	Consistent
		Reference Master data	Representation
		Root cause analysis Data Governance	Value-Added
People data	Incomplete data Outdated data	Root cause analysis	Accuracy
		Protocols for handling issues	Security
		Understanding policies	Value-Added
		Data ownerships	Timeliness
		Technology	Relevance
		Tracking issues	
Service/Asset data	Missing data, Incomplete data Incorrect values Outdated data Duplicates	Requirements	Accessibility
		Metrics	Value-added
		Measuring,	Completeness
		SLAs	Timeliness
		Policies	Interpretability
		Standards Root cause analysis Business impact analysis	
Supply Chain Management	Design errors Duplicates Master data issues	Metrics	Accessibility
		Measuring	Security
		Reporting	Timeliness
		SLAs	Accuracy
		Policies	Believability
		Protocols	Value-Added
		Root cause analysis	Completeness
		Data governance Visibility	

The *data quality management challenges* for different data domains reveal that in the case company, some domains are more advanced in data quality management considerations than others. Nevertheless, each domain has room for improvement. In data domains where data quality is being measured namely Item, and People data, the challenges involve *ensuring the validity of the data*. Validating the technical quality of the data is not a problem as such, but issues such as whether the correct data is written correctly require manual work. Also, even though there is a capability to monitor and report on data quality, old data created by using old rules seems to have a negative influence on data quality measurements. Even though, the overall capabilities for measuring data quality are not ideal, some domains completely lack data quality measuring and reporting, including service level agreements for data quality. The reaction times to data quality issues require improvement, and require protocols and processes that provide guidelines for reacting on data quality issues. Data governance and data quality policies also require improvement to

enhance their maturity. Steering data quality can be challenging, if not impossible, without data governance and defined policies. How activities are named internally also has significance, as for example governance may have negative connotation. For example, the case company utilises *data cluster steering group meeting* instead of governance for this reason. Data quality related roles and responsibilities would be beneficial to be better defined, this is even if ownerships are in place. In addition, root cause analyses and prioritising data issues are necessary to improve data quality. Some data domains also require defined data standards and business rules to act as a foundation for managing data quality.

The classifications of data dimensions according to PSP/IQ model revealed that the five *most valued data requirements* across the data domains include; *Believability, Value-added, Accessibility, Accuracy, and Security*. Believability is the extent to which data is regarded as true and credible. Value-added refers to the extent to which data is beneficial and provides advantages from its use. Accessibility denotes the extent to which data is available, or retrievable quickly and with ease. Accuracy refers to the extent of how well data values are in line with the source of correct data, and values are correct and reliable. Security, on the other hand refers to the extent to which access to data is restricted to maintain its security.

The emphasis on different data requirements varies across the data domains. Believability is particularly emphasised in those data domains where the daily work is influenced by the considerations whether data can be believed, such as Item and Company data. Challenges such as duplicates cause the believability considerations. Any data issues in customer service base to service technicians result in accessibility requirements being emphasised, typical to Service/Asset data domain. Accuracy is emphasised in domains such as people data where employee information and related tasks require accuracy and security. Security is also emphasised in data domains that involve aspects that can be seen to be important assets for the case company. Interestingly company data is not one of these, even if it contains data about customers. In supply chain management data domain, the requirements are all equally important due to many master data being deployed in the processes, all being critical elements to the order/delivery process. Requirement for value added is common to all data domains, potentially indicating that the domains have understood the impact of data quality in the business environment. Data requirements also seem to have linkages to experienced challenges aside the purpose and use of the data.

4.2 Ensuring data quality for product data management

Figure 3 illustrates the proposed model for managing data quality to enable operationalising data quality assessment and improvement. The model is a combination of theoretical and empirical findings, designed to be applicable in different processes and projects, specifically in the case company, but also by others.

The proposed model consists of seven phases, including: Data cluster steering group meeting, Policy, Data standards, Dimensions, Technology, Protocols, and Performance. The model also contains processes, sub-processes, and tools & methods. The processes include; current state analysis, discover problems & prioritise, measurement, prioritising data quality issues, renewing the information flow, remediation, and monitoring. The processes are divided into sub-processes, and presented with relevant tools & methods separately in Table 2.

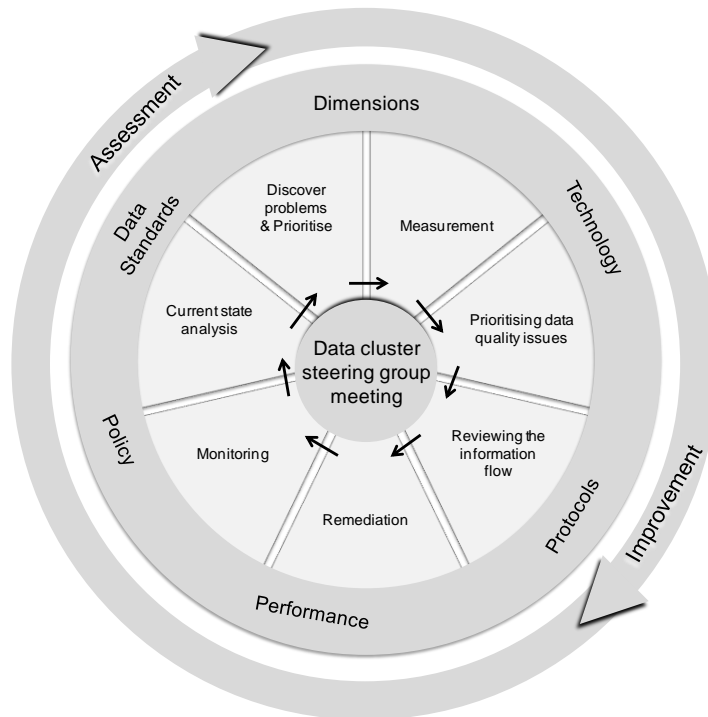


Figure 3. Model for operationalising data quality assessment and improvement.

The proposed model follows a continuous circle of assessment and improvement. The founding idea behind the circle is that all the seven phases must be possible to assess and improve. One could see the circle as a “plywood board” that cannot be bent indefinitely towards one of the corners without breaking. Hence, data quality requires all the seven phases equally, and should one of them “break”, or be missing, data quality would not be adequate. The circle can be extended should the data quality require. Nevertheless, the circle and its content provide a starting point for operationalising data quality assessment and improvement. The clockwise revolution acts as a learning process for companies.

Table 2. Data quality assessment and improvement model – processes, sub-processes, and relevant tools & methods

Process	Sub-process	Tool/ method	Purpose
Current state analysis	Information product (i.e. Data)	IP-MAP (Visualisation)	Assess the state of data quality
	Information process (Environment where data is located & influenced by processes)		Visualising data and information environment
Discover problems & Prioritise	Business impact / risk analysis	Business impact matrices Stakeholder interviews	Clarify the extent of data issues Prioritise issues impacting data quality

Measurement	Requirements Business rules Dimensions/ metrics	Interviews ETL-tools (Extract-Transform-Load) PSP/IQ (Product and Service Performance model for Information Quality)	Provide a reference point for data quality improvement
Prioritising data quality issues	Severity, impact, resolution feasibility	Prioritisation matrix	Ensure that the most relevant issues for business and operations are improved
Reviewing the information flow	Information product (i.e. Data) Information process (Environment where data is located & influenced by processes)	IP-MAP (Visualisation)	Better understand the information flow Enable finding root causes
Remediation	Prevention, Auditing, Correction, and Usage	Activity/ Data matrix ETL-tools (Extract-Transform-Load)	Addressing data issues Data correction/ cleansing
Monitoring	Issue tracking and service level agreements	ETL-tools (Extract-Transform-Load)	Enable ongoing improvement Aid in reporting on improvement

Current state analysis is conducted to assess data quality by analysing the current information product and processes. This can be done for different data domains, or processes to emphasise the perspectives of different data uses. As data and the environment where it is located have somewhat abstract nature, visualisation in some form may prove beneficial. Information Product Map (IP-MAP) method for visualisation can aid in revealing data dependencies and inconsistencies of different systems that are affecting the data, or affected by the data. The visualisation can also aid in revealing data related challenges.

Discover problems and prioritise aims to further clarify the extent of data issues by analysing how they will influence business and operations. This is to discover true challenges and prioritise issues. Business impact and risk analyses can provide means to revealing the extent of data issues. The methods can include involving relevant stakeholders and interviewing key users and specialists to discover and obtain deeper understanding on data quality issues. Utilising a business impact matrix can assist in prioritising business critical issues in early stages of the assessment and improvement efforts. The matrix can help in justifying and communicating data quality improvement needs to the management by highlighting different data related issues and their impacts.

Measurement is a prerequisite for improving data quality. Measurement can tackle the challenge of measuring data quality in different data domains. Similarly as physical products, also information products have different quality dimensions. Nevertheless, data is usually used by multiple users and is not reduced similarly as materials along their use. Dimensions exemplify different aspects of data quality that can be measured, those that can be utilised to observe data quality. Metrics, on the other hand, help in quantifying data quality assessment. It is vital that metrics used for measuring data quality are based on relevant requirements and match with the expectations. The information systems, user

expectations and business rules should be acknowledged when defining the metrics. The methods can include identifying and interviewing relevant users, utilising the PSP/IQ model for assessing the importance of data quality dimensions. Business rules enable focusing on the right aspects of data quality. Automated profiling and Extract-Transfer-Load tools can be used to perform measurements. However, it may be wise to confirm the validity of measurements by conducting occasional key user interviews.

Prioritising data quality issues ensures that those issues that are the most relevant for business and operations are improved. Prioritising also helps in avoiding wasting resources and unnecessary work. In addition, the scope of the improvement efforts can be influenced during the prioritisation. Also, should there be any needs to get commitment, or justify the data quality efforts further, prioritising can provide some support. Naturally, proper prioritisation can occur only once the uncared data has been revealed by measuring data quality against valid rules. Prioritisation must be based on the severity of data issues and their impact on business and operations, not to forget the feasibility of addressing them. Prioritisation matrices may prove beneficial for comparing data quality issues.

Reviewing the information flow attempts to analyse the flow of pieces of data from their sources to the information product, and within the information process. Reviewing and updating the IP-MAP made during current state analysis is one method that may help in analysing the information flow and the behaviour of different issues. The updated information product mapping can also assist in identifying the introduction points for data issues and challenges. An experienced practitioner can discover data quality challenges and provide insights into next steps in the improvement process.

The *remediation* phase includes addressing data issues by the means of the traditional data correction/cleansing. Options for remediation involve preventing, auditing, correcting, or accepting the current data condition. Auditing can help in evaluating and reporting data set compliance according to data quality expectations. During remediation, it is important to keep in mind that data quality is dependent on the actual data usage, hence the needs of different data domains must be appreciated. Different matrices can assist in data quality remediation efforts. Also, the ETL-tools provide many effective data quality correction features for ongoing, and one-time batch corrections and the practical remediation work. Training personnel may be needed in case of operational improvements.

Monitoring data quality is needed to enable ongoing improvements. Continuous monitoring, creating a system for tracking issues can help practitioners in reporting on the data quality improvement efforts. Monitoring must appreciate the needs of business and operations. Service level agreements may be necessary when reaching towards higher maturity of data quality management. Different stakeholders can have defined, expected data quality levels for different data usages.

5 Conclusions

Data quality has significance to companies as data forms the origins of products through key business objects, similarly as human DNA forms genetics for people. Poor data quality can have severe negative impacts on companies. However, many aspects of data quality make it a challenging concept to operationalise. This study approaches data quality from the perspective of operationalisation.

The data quality challenges in the case company were analysed by those data domains that have master data characteristics and involve multiple processes, including item data, company data, people data, service/asset data and supply chain management. Item data was discussed in more detail to exemplify the chain of logic. As Item data typically involves

the highest level of complexity, the applicability to other domains is well ensured. Both, differences and similarities were identified in the challenges experienced by different data domains. Overall, the most common data challenges in the case company across the data domains involve duplicates, irrelevant, missing, or outdated data. Thus, the consequences of poor data quality seem to have some dependence on the data domain, and the life cycle stage of the product. Highlighting the varying emphasis of data requirements across the data domains and addressing challenges accordingly is an important contribution. The differences in capabilities of data domains have an impact on data quality management. Classifications according to the analysed data dimensions revealed that the five most valued data requirements across the data domains include; Believability, Value-added, Accessibility, Accuracy, and Security. However, as the emphasis of data requirements varies across the data domains, once addressing data requirements and aiming to improve data quality, each requirement can be so significant that they may require their own project. These specific projects are able to advance data quality through specific training and definitions, reaching the level of practical data ownership.

Data quality is best ensured when organisation specific aspects are taken into account. Acknowledging varying emphasis may help in better understanding organisation specifics. Without adequate data quality, and fitness for use, product data management risks losing its significance. The model for managing data quality, proposed in this study can provide a starting point for operationalising data quality assessment and improvement. The proposed model consists of seven phases of Data cluster steering group meeting, Policy, Data standards, Dimensions, Technology, Protocols, and Performance; and contains relevant processes, sub-processes, tools and methods. The model follows a continuous circle of assessment and improvement, whilst the founding idea is that all the seven phases must be possible to assess and improve, whilst they all have a clear purpose. Comprehending the importance of data quality and purpose of activities can improve commitment within a company. The circle can be seen as a “plywood board” that cannot be bent indefinitely towards one of the corners without breaking. Hence, effective data quality requires all the seven phases equally. The circle can be extended or modified based on data quality needs. The model may prove the most effective when the needs of data domains or processes are emphasised from the perspectives of different data uses.

The practical implications of this study include highlighting the importance of organisation specific aspects of data quality and their significance on product data management. Companies may find it difficult to move forward in data quality considerations, or even get started. This difficulty is caused by different matters being linked and the understanding at different organisational levels potentially lacking. These levels include management, middle management, and those involved in practical work, whilst relevant challenges can exist on one, or at multiple levels. It is important to recognise where the potential challenges reside to address them effectively. Hence, this study, and the proposed model can act as a support for organisational learning, or as a competence gap type of analysis. As a result, the areas that are lacking can be considered from the human resources perspective to develop those areas. Naturally, one needs to first comprehend that there is a gap in expertise. Considering a *triangle of data, processes and IT* and understanding that there needs to be a balance among the three corners is a good starting point for data quality assessment and improvement. Improving the capabilities of existing personnel may prove beneficial as outsiders may have difficulties in truly having an access to the “origins of products” due to their importance. Change management and increasing personnel competence at different organisational levels can offer possibilities to get started with the presented model, or to speed up the revolutions around the data quality assessment and improvement circle. Change can only occur if it can be measured. For

example, measuring data quality enables noticing change. Concretising the core essence of products in the form data is significant for companies. Truly appreciating the importance of quality data improves data maintenance processes, IT solutions, Data quality and expertise, all these form the basis for handling the origins of products.

The scientific implications of this study involve providing means to address data quality issues more domain, or process specifically rather than by simply following the previous models that typically have a generic high-level approach. The model attempts to acknowledge the needs of different organisational levels, whilst data, processes and IT are appreciated as three cornerstones for data quality assessment and improvement. In addition, involving methods and tools for different steps are seen to potentially help in operationalising data quality. This work contributes towards the previous discussion in the literature by providing one example of taking organisation and domain specific aspects into account. This provides an input to the discussion on data quality acknowledging the significance of the surrounding setting.

The limitations of this study include analysing the practices of a single, yet significant company. Even though the proposed model has elements that can be adopted by others, it has been designed specifically for the needs of the case company. Future research could assess the applicability of the model in multiple companies to gain further experience and further test its true generalisability.

References

- Alpar, P. and Winkelsträter, S. (2014) "Assessment of data quality in accounting data with association rules", *Expert Systems with Applications*, Vol. 41, No. 5, pp. 2259–2268.
- Arlbjørn, J.S., Wong, C.Y. and Seerup, S. (2007) "Achieving competitiveness through supply chain integration", *International Journal of Integrated Supply Management*, Vol. 3 No. 1, pp. 4–24.
- Ballou, D.P. and Pazer, H.L. (1985) "Modeling data and process quality in multi-input, multi-output information systems", *Management Science*, Vol. 31, No.2, pp. 150–162.
- Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009) "Methodologies for data quality assessment and improvement". *ACM Computing Surveys*, Vol. 41, No. 3, Article 16.
- Batini, C., Barone, D., Mastrella, M., Maurino, A. and Ruffini, C. (2007) "A Framework and a Methodology for Data Quality Assessment and Monitoring", *ICIQ*, pp. 333–346.
- Blake, R. and Mangiameli, P. (2011) "The effects and interactions of data quality and problem complexity on classification", *Journal of Data and Information Quality*, Vol. 2, No. 2, Article 8.
- Breur, T. (2009) "Data quality is everyone's business designing quality into your data warehouse part 1", *Journal of Direct, Data and Digital Marketing Practice*, Vol. 11, No. 1, pp.20–29.
- Caballero, I., Caro, A., Calero, C. and Piattini, M. (2008) "IQM3: Information quality management maturity model", *Journal of Universal Computer Science*, Vol. 14, No. 22, pp. 3658–3685.
- Caballero, I., Gomez, O. and Piattini, M. (2004) "Getting better information quality by assessing and improving information", Proceedings of the ninth international conference on information quality, pp. 44–58.
- Chuang, S.-H. and Lin, H.-N. (2013) "The roles of infrastructure capability and customer orientation in enhancing customer-information quality in CRM systems: Empirical evidence from Taiwan", *International Journal of Information Management*, Vol. 33, No. 2, pp. 271–281.
- English, L.P. (2003) "Total information quality management: A complete methodology for IQ management", *Dm Review*, Vol. 9, No. 3.
- Eppler, M.J. (2006) "Managing information quality: increasing the value of information in knowledge-intensive products and processes. 2nd ed. Springer, Berlin, Germany.

- Eppler, M. and Helfert, M. (2004) "A classification and analysis of data quality costs", MIT International Conference on Information Quality, November 5-6, 2004, Boston, MA.
- Eppler, M.J. and Wittig, D. (2000) "Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years", Proceedings of the Information Quality Conference, University of St. Gallen, 2000, p. 83-96.
- Ferreira, F., Coimbra, M.E., Bairrão, R., Viera, R., Freitas, A.T., Russo, L.M.S. and Borbinha, J. (2014) "Data Management in Metagenomics: A Risk Management Approach", *International Journal of Digital Curation*, Vol. 9, No. 1, pp. 41-56.
- Glowalla, P. and Sunyaev, A. (2014) "ERP system fit – an explorative task and data quality perspective", *Journal of Enterprise Information Management*, Vol. 27, No. 5, pp.668-686.
- Gustafsson, P., Lindstrom, A., Jagerlind C. and Tsoi, J. (2006) "A Framework for Assessing Data Quality-from a Business Perspective", *Software Engineering Research and Practice*, pp. 1
- Haug, A., Zachariassen, F. and van Liempd, D. (2011) "The costs of poor data quality", *Journal of Industrial Engineering and Management*, Vol. 4, No. 2, pp. 168-193.
- Haug, A., Arlbjørn, J.S., Zachariassen, F. and Schlichter, J. (2013) "Master data quality barriers: an empirical investigation", *Industrial Management and Data Systems*, Vol. 113, No. 2, pp. 234-249.
- Haug, A. and Arlbjørn, J.S. (2011) "Barriers to master data quality", *Journal of Enterprise Information Management*, Vol. 24, No. 3, pp. 288-303.
- Hazen, B.T. Boone, C.A., Ezell, J.D. and Jones-Farmer, L.A. (2014) "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications", *International Journal of Production Economics*, Vol. 154, pp. 72–80.
- Hüner, K.M., Schierning, A., Otto, B. and Österle, H. (2011) "Product data quality in supply chains: the case of Beiersdorf", *Electronic Markets*, Vol. 21, No. 2, pp. 141-154.
- Hüner, K.M., Ofner, M. and Otto, B. (2009) "Towards a maturity model for corporate data quality management", Proceedings of the 2009 ACM symposium on Applied Computing, pp. 231-238
- 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.
- Kahn, B.K., Strong, D.M. and Wang, R.Y. (2002) "Information quality benchmarks: product and service performance", *Communications of the ACM*, Vol. 45, No.4, pp. 184-192.
- Khatri, V. and Brown, C.V. (2010) "Designing data governance", *Communications of the ACM*, Vol. 53, No. 1, pp. 148-152.
- Khosroanjom, D., Ahmadzade, M., Niknafs, A. and Mavi, R.K. (2011) "Using fuzzy AHP for evaluating the dimensions of data quality", *Int. J. of Business Information Systems*, Vol.8, No.3, pp.269-285.
- Kulikowski, J.L. (2014) "Data Quality Assessment: Problems and Methods", *International Journal of Organizational and Collective Intelligence*, Vol. 4, No. 1, pp. 24-36.
- Kwon, O., Lee, N. and Shin, B. (2014) "Data quality management, data usage experience and acquisition intention of big data analytics", *International Journal of Information Management*, Vol. 34, No. 3, pp. 387–394.
- Leung, H.K.N. (2001) "Quality metrics for intranet applications", *Information & Management*, Vol. 38, No. 3, pp. 137-152.
- Loshin, D. (2011) "The Practitioner's Guide to Data Quality Improvement", Elsevier, New York.
- Madnick, S.E., Wang, R.Y., Lee, Y.W, Zhu, H. (2009) "Overview and Framework for Data and Information Quality Research", *Journal of Data and Information Quality*, Volume 1 Issue 1, Article No. 2.
- Merton, R., Fiske, M. and Kendall, P. (1990) "The Focused Interview: A Manual of Problems and Procedures", 2nd ed., The Free Press, New York, NY.

- Miksa, T., Strodl, S. and Rauber, A. (2014) "Process Management Plans", *International Journal of Digital Curation*, Vol. 9, No. 1, pp. 83-97.
- Motiwalla, L. and Li, X.-B. (2013) "Developing privacy solutions for sharing and analysing healthcare data", *Int. J. of Business Information Systems*, Vol.13, No.2, pp.199-216.
- O'Donoghue, J., O'Kane, T., Gallagher, J., Courtney, G., Aftab, A., Casey, A., Torres, J. and Angove, P. (2011) "Modified Early Warning Scorecard: The Role of Data/Information Quality within the Decision Making Process", *Electronic Journal of Information Systems Evaluation*, Vol.14, No. 1, pp. 100-109.
- Ofner, M.H., Straub, K., Otto, B. and Oesterle, H. (2013) "Management of the master data lifecycle: a framework for analysis", *Journal of Enterprise Information Management*, Vol. 26, No. 4, pp.472-491.
- Ofner, M.H., Otto, B. and Österle, H. (2012) "Integrating a data quality perspective into business process management", *Business Process Management Journal*, Vol. 18, No. 6, pp. 1036-1067.
- Orr, K. (1998) "Data quality and systems theory", *Communications of the ACM*, Vol. 41, No. 2, pp. 66-71.
- Otto, B. (2012a) "Managing the business benefits of product data management: the case of Festo", *Journal of Enterprise Information Management*, Vol. 25, No. 3, pp. 272-297.
- Otto, B. (2012b) "How to design the master data architecture: Findings from a case study at Bosch", *International Journal of Information Management*, Vol. 32, No. 4, pp. 337-346.
- Otto, B., Wende, K., Schmidt, A. and Osl, P. (2007) "Towards a framework for corporate data quality management", *ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems*, pp. 916-926.
- Paulk, M.C., Curtis, B., Chrissis, M.B. and Weber, C.V- (1993) "Capability maturity model, version 1.1. *IEEE Software*, Vol. 10, No. 4, pp. 18-27.
- Puurunen, A., Majava, J. and Kess, P. (2014) "Exploring incomplete information in maintenance materials inventory optimization", *Industrial Management & Data Systems*, Vol. 114, No. 1, pp. 144-158.
- Redman, T.C. (1998) "The impact of poor data quality on the typical enterprise". *Communications of the ACM* 41(2): 79-82.
- Rocha, A. (2012) "Three-dimensional model for the global quality of a website", *International Journal of Business Information Systems*, Vol. 10, No. 4, pp. 436-446.
- Ryu, K.S., Park, J-S., Park, J.H. (2006), "A data quality management maturity model", *ETRI Journal*, Vol. 28 No. 2, pp. 191-204.
- Shahin, A., Pool, J.K. and Poormostafa, M. (2014) "Evaluating and ranking hotels offering e-service by integrated approach of Webqual and fuzzy AHP", *International Journal of Business Information Systems*, Vol. 15, No. 1, pp. 84-104.
- Shankaranarayanan, G. and Cai, Y. (2006) "Supporting data quality management in decision-making", *Decision Support Systems*, Vol. 42, No. 1, pp. 302-317.
- Silvola, R., Jaaskelainen, O., Kropsu-Vehkaperä, H., Haapasalo, H. (2011), "Managing one master data – challenges and preconditions", *Industrial Management & Data Systems*, Vol. 111 No.1, pp.146-62.
- 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.
- Smith, A.D. (2011) "Quality assurance practices for competitive data warehouse management systems", *Int. J. of Business Information Systems*, Vol.7, No.4, pp.440-457.
- Sundaraman, A. (2011) "A framework for linking Data Quality to business objectives in decision support systems," *Trendz in Information Sciences and Computing*, pp.177-181.
- Tayi, G.K. and Ballou, D.P. (1998) "Examining Data Quality", *Communications of the ACM*, Vol. 41, No. 2, pp. 54-57.

- Van Zeist, R.H.J. and Hendriks, P.R.H. (1996) "Specifying software quality with the extended ISO model", *Software Quality Journal*, Vol. 5, No. 4, pp. 273-284.
- 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 the ACM*, Vol. 41, No. 2, pp. 58-65.
- Wang, R.W. and Strong, D.M. (1996) "Beyond Accuracy: What Data Quality Means to Data Consumers", *Journal of Management Information Systems*, Vol. 12, No. 4, pp. 5-33.
- Weber, K. Otto, B. and Österle, H. (2009) "One Size Does Not Fit All - A Contingency Approach to Data Governance", *Journal of Data and Information Quality*, Vol. 1, No. 1, Article 4.
- Weill, P. and Ross, J. (2004) *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results*, Boston: Harvard Business School Press.
- Wende, K. (2007) "A model for data governance - Organising accountabilities for data quality management", ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems, pp. 416-425.
- Woodall, P., Borek, A. and Parlikad, A.K. (2013) "Data quality assessment: The Hybrid Approach", *Information and Management*, Vol. 50, No. 7, pp. 369-382.
- Xiao, Y., Lu, L.Y.Y., Liu, J.S. and Zhou, Z. (2014) "Knowledge diffusion path analysis of data quality literature: A main path analysis", *Journal of Informetrics*, Vol. 8, No. 3, pp. 594-605.