Theoretical Support for Enhancing Data Quality: Application in Electronic Medical Records

Caihua Liu
University of Technology Sydney
Caihua.Liu@student.uts.edu.au

Amir Talaei-Khoei
University of Nevada, Reno
University of Technology Sydney
atalaeikhoei@unr.edu

Didar Zowghi
University of Technology Sydney
Didar.Zowghi@uts.edu.au

Abstract

This paper aims at reviewing the existing theoretical support to enhance data quality and utilizing the findings of the review in the context of electronic medical records (EMRs). For this to happen, we first conducted a survey of publications that have a focus on an empirical investigation of factors influencing data quality in the conceptual models. By using a well-established taxonomy development method from the discipline of information systems, we then proposed 3 dimensions for studying factors influencing data quality and constructing the conceptual model for enhancing data quality: breadth, depth, and interaction, within 9 characteristics under different dimensions. Last, we compared related studies using the proposed dimensions and utilized the findings of the review in enhancing EMRs quality to disclose the limitations and possibilities of new areas for further study.

Keywords

Enhance data quality, review of theoretical support, electronic medical records.

Introduction

Most industries rely on their data for operation, decision making and planning. The availability of large volumes of relevant and useful data could help organizations become more productive and competitive (Shaw et al. 2001). High-quality data assists organizations in working efficiently; while bad data costs the economy trillions of dollars a year (O’Brien 2018). Undoubtedly, poor data quality can have substantial social and economic impacts on organizations (Wang and Strong 1996). For instance, poor-quality data can have negative effects on customer satisfaction, decision-making processes, services performance, and employee job satisfaction (Kahn et al. 2002). At the same time, organizations increase operational costs on detecting and correcting data errors and business users may lose confidence of any initiatives based on such data (Haug et al. 2011).

To enhance data quality, organizations and individuals need to know what impacts data quality. If the factors influencing data quality and all the possible relationships between these factors are uncovered, data quality would be potentially preserved or systematically improved. Although prior studies have summarized possible factors influencing data quality (Nord et al. 2005; Xu 2013; Xu et al. 2002), they failed to identify the relationships among these factors. Some researchers have constructed conceptual models of managing and improving data quality with hypotheses of the relationships among factors affecting data quality (Al-Hiyari et al. 2013; Tee et al. 2007; Xiao et al. 2009). However, they did not compare the differences between related studies. Therefore, the structural characteristics of conceptual models for factors influencing data quality are not available in the existing literature. A taxonomy can provide a means to organize and structure
knowledge of a field, thus enabling researchers to study relationships among concepts (Glass and Vessey 1995; McKnight and Chervany 2001). Accordingly, a taxonomy is a promising tool to order factors influencing data quality in order to improve our understanding about differences in related studies and constructs of the conceptual model for enhancing data quality.

In this paper, we aim to examine the theoretical support for enhancing data quality. We present our attempt in organizing and structuring the characteristics of prior work on factors influencing data quality in order to give insights for further study on enhancing data quality. To achieve our aims, we first conduct a preliminary survey on the conceptual models of factors affecting data quality and then review the characteristics of these models. Although the paper does not provide an empirical result, three dimensions developed in this paper using the method of taxonomy development (Nickerson et al. 2013) can be utilized to compare related studies, providing support for constructing the conceptual model to enhance data quality.

We conceptualize an initial taxonomy using relevant theories and use it to study and analyze factors affecting data quality identified in the conceptual models through the review of research papers selected based on our criteria. At each step of taxonomy development, we test our ending conditions and revise the taxonomy accordingly. Using this taxonomy development method, we propose three dimensions for related studies on factors influencing data quality: breadth, depth, and interaction within a set of characteristics under different dimensions.

This paper also illustrates our findings of the review in the context of Electronic Medical Records (EMRs) as an exemplar because the widespread use of EMRs and data quality problems are more prominent in healthcare but have not been taken seriously in practice (Christopher 2014). As an example of data quality problems in EMRs that motivated this illustration, we consider the case of a system developed in 2015, aimed at improving diabetes management for patients across thirty-six medical clinics in Australia, through the continuous analysis of daily-generated EMRs. An analytics infrastructure was designed and implemented to link EMRs for diabetes patients across these thirty-six clinics, and to automatically collect the daily-generated data from the EMRs database on a weekly basis. The EMRs were de-identified but linked through a secure-ID generation platform. The system also provided an application that was available on the physicians’ office computer, and could predict the patients at risk of developing diabetes. The application used statistical forecasting based on similarities of patient data with those who had diabetes, and was able to differentiate between types of diabetes. Despite its potential for improving the quality of diabetes disease management, the project failed to deliver its objectives. The reason for the failure was traced to the poor quality of data entry in the EMRs in clinics that led to ineffective EMRs quality. Therefore, our study suggests specific concepts for enhancing EMRs quality.

In the next section, we give a brief introduction about data quality. We then present our research methods to review related studies. Thereafter, we summarize the characteristics of prior work and utilize the findings of the review within the context of EMRs. In the last section, we conclude the paper, and discuss the limitations and future work.

What is Data Quality?

The definition of quality, as noted in International Standards Organization (ISO) 8402: 1986, was described as “the totality of features and characteristics of an entity that bears on its ability to satisfy stated and implied needs” (ISO 1986). Similarly, the concept of quality was defined as “fitness-to-purpose” using the theory of Total Quality Management (Klobas 1995; Ziegel 1990). Based on these theories, data quality is consistently defined in terms of its “fitness-to-purpose” (Klobas 1995; Ziegel 1990). Because definitions of data quality address both ‘fitness-for-purpose’ aspect and ‘ability to satisfy needs’ aspect, operational definitions of data quality are context specific (Fehrenbacher and Helfert 2012). Nevertheless, in the information systems tradition, researchers have grouped definitions of data quality into dimensions and proposed measures for data quality (Wang and Strong 1996). For example, scholars defined several dimensions of data quality such as completeness, currency, and consistency that achieve context specific nature of data quality (Ji-fan Ren et al. 2017; Miller 1996; Wixom and Todd 2005).

The literature on information quality and data quality has developed in parallel, however, there have been attempts to draw a distinction between the two from the definitions of data and information. In the discipline of information systems, data can be defined as symbols that are stored in a database and manufactured and used by an information system may describe facts in the real world (English 1999; Price
and Shanks 2005). Information refers to the data that is processed and interpreted by an information system, giving a meaning to a human user (English 1999; Glowalla and Sunyaev 2014; Price and Shanks 2005). Tilly et al. (2017) clearly indicated that data is objective to present a phenomenon unconcerned the information system while information is subjective to put the data into context using the information system that users can understand. In this work, we are interested in the quality of the data that is stored in the database and processed by the information systems, and therefore we focus on data quality.

Research Methods

Selection of Research Papers

Before conceptualizing the characteristics of related studies, we conducted a preliminary survey on the conceptual models of factors influencing data quality. We used Google Scholar as a general database in order to cover a broad range of disciplines and adopted the following search query for our search: (factor OR impact OR influence OR affect OR determinant) AND “data quality” AND (taxonomy OR classification OR category OR typology OR model OR framework). Furthermore, we screened the papers published in ICIS, AMCIS, ECIS, PACIS and HICSS proceedings. We finally identified 17 related studies during the period 2000-2017. We selected 10 papers out of surveyed studies in this work for further study based on the following criteria: (1) the studies have a focus on an empirical investigation of factors influencing data quality; (2) the factors derived from the studies have been conceptualized in the conceptual models; and (3) the studies have been published in conference, journal or book chapter (because Google Scholar includes non-published work as well).

Conceptualization of Characteristics of Related Studies

To build and study the characteristics of the conceptual models on factors influencing data quality, we apply an iterative process for taxonomy development as suggested by Nickerson et al. (2013). This method of taxonomy development from information systems discipline enables us to organize and structure the characteristics of related studies using an established, rigorous procedure for understanding the conceptual models of factors influencing data quality.

Definitions: Dimension, Characteristic, and Meta-characteristic

Taxonomy development can be done either based on a single dimension or based on a number of dimensions (Bailey 1994). Dimension is generally categorical data and can be also called as variable in the taxonomy (Nickerson et al. 2013). In this paper, we use the term dimension to describe a category of characteristics about factors affecting data quality identified and conceptualized in the conceptual models. As mentioned in Bailey (1994)’s foundational book on classification techniques, characteristic is used to describe the fundamentals of the phenomenon. Accordingly, we define characteristic as a subcategory of factors affecting data quality identified and conceptualized in the conceptual models, being served to delineate an aspect of the dimension on the taxonomy. For meta-characteristic, as Nickerson et al. (2013) notes, it is used as the basis for the choice of characteristics in the taxonomy. We describe the meta-characteristic as root causes of data quality problems.

With these explicit definitions in mind, we then follow the method of taxonomy development (Nickerson et al. 2013) to review the conceptual models of factors influencing data quality as summarized below.

Taxonomy Development

Step 1: Determine meta-characteristic

The objects of interest in our study are specific factors influencing data quality identified and conceptualized in the conceptual models. We have defined the meta-characteristic of objects of interest for classification as root causes that result in data quality problems, assisting in the tasks of studying the characteristics of related studies and grouping the factors influencing data quality.

Step 2: Determine ending conditions
We will end the taxonomy development when the taxonomy meets both objective and subjective conditions. The objective conditions mean that each dimension in the taxonomy approaches “mutually exclusive and collectively exhaustive characteristics” (Nickerson et al. 2013), while the subjective conditions concern whether a taxonomy achieves concise, robust, comprehensive, extendible, and explanatory criteria based on the researcher’s viewpoints (Nickerson et al. 2013). In this study, we adapt and develop three objective ending conditions and five subjective conditions (Nickerson et al. 2013) for determining the end of our taxonomy development. See Table 1.

<table>
<thead>
<tr>
<th>Ending conditions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>No new dimensions or characteristics could be added in the last iteration. Every dimension is unique and not repeated. No more factors influencing data quality in the conceptual models need to be examined.</td>
</tr>
<tr>
<td>Subjective</td>
<td>Concise</td>
</tr>
<tr>
<td></td>
<td>Robust</td>
</tr>
<tr>
<td></td>
<td>Comprehensive</td>
</tr>
<tr>
<td></td>
<td>Extendible</td>
</tr>
<tr>
<td></td>
<td>Explanatory</td>
</tr>
</tbody>
</table>

Table 1. Ending Conditions Used in Our Taxonomy Development

**Step 3: Conceptualize characteristics and dimensions of the taxonomy**

First, because information products manufactured and used in the enterprise management process can be viewed as the products generated within a processing system on raw data, the theory of product quality management can be utilized to study data quality (Wang 1998). In product quality management, factors influencing product quality are divided into three groups: human, management and technology (Xiao et al. 2009). Therefore, we can view the root causes that result in data quality problems from human, managerial or technical perspective. Additionally, many researchers revealed that organizational and external factors have impacts on data quality (Nord et al. 2005; Xu 2013; Xu et al. 2002). Hence a holistic list of factors influencing data quality contains five main groups: human, organizational, managerial, technical, and external factors (Liu et al. 2018). These theories then guide us to determine and classify the factors influencing data quality in the conceptual models. Essentially, these five groups also address the characteristics of related studies from a breadth perspective that presents the extent to which factors affecting data quality are included in the conceptual models. We now have a dimension breadth to group human, organizational, managerial, technical and external factors influencing data quality in the taxonomy.

Second, related studies (e.g. Nord et al. (2005), Xu (2013) and Xu et al. (2002)) implies that the root causes of data quality problems can be studied from staff members perspective (individual level) and/or top management perspective (organizational level). This also assist us view the factors influencing data quality from a depth perspective and therefore, a dimension depth can be used to group the characteristics of related studies on factors influencing data quality into individual and organizational level.

Third, academics proposed their conceptual models of factors influencing data quality and validated the hypotheses for: (1) relationships between the factors and data quality, and (2) relationships between these factors (Al-Hiyari et al. 2013; Kokemueller 2011; Tee et al. 2007; Xiao et al. 2009). The establishment of links between the factors influencing data quality could help us enhance data quality. Accordingly, we develop a dimension interaction to describe relationships with data quality and between the factors affecting data quality in the conceptual models.

**Step 4: Examine sample papers for these characteristics and dimensions**

We review the factors influencing data quality from the selected papers and group these factors into the corresponding characteristics.

**Step 5: Create (revise) the taxonomy**

We group the characteristics into three dimensions to create our taxonomy for related studies:
• Breadth dimension: Human, Organizational, Managerial, Technical, and External Factors characteristics
• Depth dimension: Individual Level and Organizational Level characteristics
• Interaction dimension: With Data Quality and Between the Factors characteristics

Step 6: Determine whether meet ending conditions

We have examined all the factors affecting data quality that are identified in our sample papers. No more objects of interest can be examined. Additionally, we have added no new dimensions and each dimension is unique and not repeated. Accordingly, we meet the objective ending conditions. In terms of subjective ending conditions, we consider that our taxonomy achieves concise, extendible, comprehensive, robust and explanatory criteria (as listed in Table 1) and therefore, our taxonomy development ends.

Comparison of Theoretical Support for Enhancing Data Quality

Table 2 summarizes the characteristics of the conceptual models on factors influencing data quality, including 3 dimensions with 9 individual characteristics that are grouped under different dimensions.

<table>
<thead>
<tr>
<th>Author(s)/References</th>
<th>Source</th>
<th>Number of Citations</th>
<th>Field of Application</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Breadth</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HF OF MF TF EF INDL ORGL With Data Quality Between the Factors</td>
</tr>
<tr>
<td>Wixom and Watson</td>
<td>MIS Quarterly</td>
<td>1524</td>
<td>Multiple industries</td>
<td>X X X X X X X</td>
</tr>
<tr>
<td>(2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xu et al. (2002)</td>
<td>ITBM: Challenges and Solutions</td>
<td>9</td>
<td>Multiple industries</td>
<td>X X X X X X X</td>
</tr>
<tr>
<td>Nord et al. (2005)</td>
<td>JDM</td>
<td>25</td>
<td>X</td>
<td>Multiple industries</td>
</tr>
<tr>
<td>Tee et al. (2007)</td>
<td>Accounting &amp; Finance</td>
<td>44</td>
<td>A single case study organization</td>
<td>X X X X</td>
</tr>
<tr>
<td>Xia et al. (2009)</td>
<td>ICMSE</td>
<td>6</td>
<td>Multiple industries</td>
<td>X X X X X X</td>
</tr>
<tr>
<td>Kokemueeller (2011)</td>
<td>AMCIS</td>
<td>4</td>
<td>X</td>
<td>Multiple industries</td>
</tr>
<tr>
<td>Xu (2013)</td>
<td>AMCIS</td>
<td>6</td>
<td>X</td>
<td>Multiple industries</td>
</tr>
<tr>
<td>Al-Hiyari et al. (2013)</td>
<td>AJE</td>
<td>23</td>
<td>Perception of students</td>
<td>X X X X</td>
</tr>
<tr>
<td>Nicol et al. (2013)</td>
<td>MedInfo</td>
<td>20</td>
<td>X</td>
<td>Healthcare</td>
</tr>
<tr>
<td>Coleman et al. (2015)</td>
<td>BMC Family Practice</td>
<td>21</td>
<td>Healthcare</td>
<td>X X X</td>
</tr>
</tbody>
</table>

HF: Human factors affecting data quality; OF: Organizational factors affecting data quality; MF: Managerial factors affecting data quality; TF: Technical factors affecting data quality; EF: External factors affecting data quality; INDL: The model addresses factors affecting data quality at an individual level; ORGL: The model addresses factors affecting data quality at an organizational level.

Table 2. Characteristics of Related Studies on Factors Influencing Data Quality

Several observations can be made in reference to Table 2. First, in the recent years the main contributions have been made to identify the root causes of data quality problems from multiple industries and various data users. Second, we note that researchers have established an empirical body of knowledge on the factors affecting data quality from breadth, depth, and interaction dimensions. For breadth dimension, the most frequently used factors that impact data quality concern human, managerial, and technical factors. With
respect to depth dimension, the factors affecting data quality both at individual and organizational levels have received significant attention. In terms of interaction dimension, related studies have empirically examined the relationships between data quality and the factors affecting data quality. Moreover, some of them have tested the relationships between the factors that impact data quality, providing great details about the phenomenon under consideration.

**Breadth Dimension**

*Breadth* dimension refers to the extent to which factors affecting data quality were included in the conceptual models, containing human, organizational, managerial, technical, and external factors characteristics.

The human factors refer to individual cognition and capability on tasks related to data quality management and improvement. The understanding of importance of data quality and perceived usefulness and need for data quality could facilitate an organization that focuses on managing data and preserving data quality (Tee et al., 2007). Furthermore, if staff members satisfy their career development aspiration, they are likely to play their roles in achieving quality data in the organization, thus following the rules and procedures when using the system and addressing the recording and/or reporting tasks carefully (Nord et al. 2005; Xu 2013; Xu et al. 2002). Additionally, personal ability determines the extent to which the tasks related to data quality management and improvement can be completed and therefore, individual competency plays an important role in enhancing data quality (Nicol et al. 2013; Xiao et al. 2009).

The organizational factors involve organizational attributes or properties (e.g. structure, location, size and industry) that could have impacts on enhancing data quality in an organization. For example, a centralized structure in an organization could conduct better controls for data quality and further obtain high-quality data (Kokemueller 2011; Xu et al. 2002). An organization that has a culture of focusing on data quality could have a better chance to achieve quality data, because data quality management and improvement could receive top management group's support, sufficient resources, and staff's participation (Nord et al. 2005; Xu 2013; Xu et al. 2002).

The managerial factors stress the organization and coordination of activities to achieve defined goals of data quality in an organization. Firstly, appropriate allocation of human resources, time and funding could guarantee effective technologies being used in the implementation of innovative information systems and data quality initiatives (Kokemueller 2011; Wixom and Watson 2001). Without management support for resources, these initiatives are unlikely to succeed. Secondly, data quality management and process management could regulate business processes in creating and using quality data and monitor staff participation in achieving quality data, thus further contributing to the enhancement of data quality in organizations.

The technical factors concern information systems and supporting technologies to support data quality management and improvement. The quality of source systems can have a profound effect on the initiatives for enhancing data quality, because standardized data can lead to easier data manipulation and further high-quality data in these initiatives (Coleman et al. 2015; Kokemueller 2011; Xu et al. 2002; Wixom and Watson 2001).

The external factors emphasize forces outside an organization and individuals that have the potential to affect data quality such as physical environment (Nord et al. 2005; Xu 2013; Xu et al. 2002). For instance, poor air-condition environment could reduce employees' work efficiency and result in human errors in the recoding and/or reporting tasks. Thus, physical environment could contribute to enhancing data quality (Nord et al. 2005; Xu 2013; Xu et al. 2002).

The model of critical success factors for data quality developed by Nord et al. (2005), Xu (2013) and Xu et al. (2002) gives a holistic picture of factors influencing data quality and approaches the greatest breadth, however, it is lacking the interactions between these factors.

**Breath Dimension in Quality of Electronic Medical Records**

From breadth dimension, related studies focused on the factors influencing data quality from human, managerial, and technical characteristic. Similarly, when studying factors affecting EMRs quality in order to achieve the quality of care, human, managerial, and technical factors should be included (Liu et al. 2018).

**Human factors influencing in EMRs quality:** User acceptance of eHealth affects an individual's attitudes and intentions towards eHealth and his or her adoption of eHealth, thus having an impact on enhancing EMRs quality. van Engen-Verheul et al. (2016) revealed that most clinical officers do not use the EMRs
much. As a result, poor-quality data may be entered into the EMRs that reduces the quality of EMRs. Additionally, healthcare practitioners violate the organizational data protocols in data management and make manual errors in data extraction that could lead to ineffective EMRs quality (Liaw et al. 2013).

**Managerial factors influencing EMRs quality:** Nowadays many clinics have decided to make the move from paper-based records to electronic records. The implementation of EMRs is an expected result, not overnight, but also not open-ended time-wise. In other words, the use of EMRs and supporting technologies is progressing at a steady rate (Jamoom et al. 2012). Thus, the clinics should improve the system of rules, standards and procedures on EMRs quality management in data transition from paper-based records to electronic records for patient safety.

**Technical factors influencing EMRs quality:** The corruption of the database architecture or information systems that could not meet users’ needs to manage routinely collected clinical data undoubtedly reduce EMRs quality (Liaw et al. 2013). In addition, Cohen et al. (2016) revealed that the limited input space of user interface could result in poor-quality data in patient records during the data entry, and van Engen-Verheul et al. (2016) also disclosed violations of alignment between the system and workflow in care processes could incur data quality problems in EMRs that have negative impacts on EMRs quality.

**Depth Dimension**

Depth dimension concerns factors influencing data quality that were studied from top down in an organization, including individual and organizational level characteristics.

The commitment to data quality management and improvement should involve both at organizational and individual level. A body of literature indicates that the commitment of top management is a key contributor to quality management (Nord et al. 2005; Xiao et al. 2009; Xu et al. 2002). Top management team (organizational level) need to make many decisions on regulations formulation, processes management and resources allocation in order to achieve the goals of data quality (Kokemueller 2011). At the same time, active engagement of staff members (individual level) plays an essential role in the success of quality management (Sharma 2015), because personal competency and attitude determines the extent to which the tasks in relation to data quality management and improvement can be completed (Nicol et al. 2013; Xiao et al. 2009). Accordingly, both individual and organizational level (also depth dimension) should be considered when studying factors affecting data quality.

**Depth Dimension in Quality of Electronic Medical Records**

The factors influencing data quality were frequently studied at both individual and organizational levels. In the context of EMRs, enhancing EMRs quality should also receive the attention from top down in clinics.

Scholars have asserted that top management commitment has a positive impact on regulation formulation and process management for the information systems implementation, because top management determines the degree to which resources can be allocated (Kokemueller 2011; Wixom and Watson 2001) and political resistance to the implementation of information systems can be dealt with (Wixom and Watson 2001). Similarly, for a clinic, the EMRs implementation together with EMRs quality management cannot ignore the support from the clinic director.

Clinical staff’s participation in EMRs context is the process whereby clinical staffs are involved in the EMRs-enabled care processes. The misunderstanding or insufficient knowledge about data entry policy could result in incomplete documentation in care processes (Kelley et al. 2015). Additionally, if a staff member lacks the awareness about importance of the recording or reporting tasks, data delays or errors might occur due to human carelessness that could result in problems associated with EMRs quality (Warsi et al. 2002). Thus, staff participation determines the degree of enhancing EMRs quality in clinics.

**Interaction Dimension**

Interaction dimension addresses relationships between data quality and the factors influencing data quality. It also covers relationships between the factors that affect data quality.

Researchers have empirically examined the relationships between the factors in their proposed model and indicated that the factors influencing data quality dynamically relate to each other (Al-Hiyari et al. 2013; Kokemueller 2011; Tee et al. 2007; Xiao et al. 2009). The underlying mechanisms of interactions between the
factors influencing data quality could help us potentially preserve quality data or systematically reduce bad data. For example, a high level of top management support is positively associated with business-IT alignment for addressing data quality problems (Xiao et al. 2009). If top management is likely to make the commitment to alignment between IT professionals and staff members and to locate required resources for data quality initiatives, the alignment between IT and business could be facilitated and further helps individuals and organizations achieve quality data. At this moment, the business-IT alignment can be considered as a mediator between top management support and data quality. We argue that factors affecting data quality cannot be studied as mutually exclusive categories, because some factors could stop and/or trigger another factor that contribute to achieving quality data. Accordingly, an establishment of links between the factors affecting data quality is a significant enabler to enhance data quality.

Interaction Dimension in Quality of Electronic Medical Records

In terms of interaction dimension, only a few studies have established the relationships between the proposed factors and data quality together with relationships between these factors (Al-Hiyari et al. 2013; Kokemueller 2011; Tee et al. 2007; Xiao et al. 2009). As we repeatedly indicated, the establishment of relationships between the factors influencing data quality should be considered before the breadth and depth dimension to enhance data quality. Accordingly, the four conceptual models developed by Al-Hiyari et al. (2013), Kokemueller (2011), Tee et al. (2007), and Xiao et al. (2009) can be viewed as candidates to suggest specific concepts for the enhancement of the EMRs quality. Meanwhile, factors influencing data quality should be studied both at individual and organizational level for data quality improvement. Among these four studies, Kokemueller (2011) and Tee et al. (2007) validated hypotheses of relationships between the factors influencing data quality proposed, however, they failed to include staff participation (at individual level) in their conceptual models. Therefore, Al-Hiyari et al. (2013) and Xiao et al. (2009) developed a stronger conceptual model for factors influencing data quality with richness and reach in breadth (approaching the most frequently used factors in breadth dimension including human, managerial and technical factors), depth, and interaction. Unfortunately, Al-Hiyari et al. (2013) utilized the data from the students’ perspective but not from industries to test their model that may impact the validity of the results. In short, Xiao et al. (2009) provided a suitable guideline within breadth, depth, and interaction of factors influencing data quality for data quality management that can be used to suggest specific concepts for enhancing the quality of EMRs.

Conclusion and Future Work

This study reviews the theoretical support available in the literature to enhance data quality. It organizes and structures the characteristics of related studies on factors influencing data quality. By using the method of taxonomy development, we have developed breadth, depth, and interaction dimensions for grouping the factors affecting data quality to construct the conceptual model of enhancing data quality. The greatest breadth dimension addresses human, organizational, managerial, technical and external factors that impact data quality (e.g. Xu (2013) and Xu et al. (2002)). The depth dimension concerns the factors influencing data quality at both organizational and individual levels. The interaction dimension addresses the establishment of relationships between the factors and data quality as well as relationships between these factors. Furthermore, we have illustrated that the dimensions proposed in this paper can be effectively utilized in the EMRs context.

For academic contributions, we have developed 3 dimensions (including breadth, depth, and interaction) and 9 characteristics under different dimensions for studying factors influencing data quality, and compared theoretical support to enhance data quality, which are lacking in the existing literature. Furthermore, we have identified that the model of factors influencing data quality (Xiao et al. 2009) provides a stronger guideline within breadth, depth, and interaction of factors influencing data quality that can be utilized to suggest specific concepts for enhancing data quality.

For practical contributions, an appreciation of this topic can be of practical use as practitioners effectively use information systems and improve their data practices. The proposed dimensions can be utilized to determine the root cause for a data quality problem and suggest a specific solution to address the problem. Because the organizational, managerial, technical and external factors could have impacts on users’ performance (human factors) in the tasks related to data quality management and improvement, the customised managerial and technical strategies could help users enhance data quality.

The present work only focuses on data quality and does not differentiate the concepts associated with information and data quality. However, there is a body of literature that distinguish these two definitions.
Theoretical Support for Enhancing Data Quality

(English 1999; Price and Shanks 2005). Researchers in this area are recommended to take the differences in definitions of information and data quality into account and redesign the present work to see the differences in the theoretical support. Furthermore, we note that researchers have investigated the factors influencing data quality at the stage of system design or at the stage of system use. Because data quality problems could occur from data creation (at the stage of system design) to its usage (at the stage of system use), we believe that IS community should also address these factors in combining both system design and system use.

This study can serve as a conceptual isobar that investigates the theoretical support to enhance data quality. Furthermore, the results presented in this paper can be used in an empirical study for enhancing EMRs quality. This work therefore encourages academics to examine the conceptual model of factors influencing data quality within the EMRs context and to develop hypotheses on the relationships between these factors through surveys and case study. Researchers could also (1) provide similar conceptual models for another application by using breadth, depth and interaction dimensions for studying factors influencing data quality, and (2) extend the prior models by adding new factors and/or new relationships between these factors by using these three criteria.

Our review has revealed that organizational and external factors influencing data quality may not have received enough attention and require further investigation. An important area of study is about the size and type of EMRs practices (organizational factors) and whether this has an impact on EMRs quality. Another topic of interest is investigating external factors and their impacts on the quality of EMRs.

REFERENCES


