Understanding Data Quality Issues in Dynamic Organisational Environments – A Literature Review

Sarah Anstiss
Olivera Marjanovic
The University of Sydney Business School
Sydney, Australia
Email: sans3482@uni.sydney.edu.au, olivera.marjanovic@sydney.edu.au

Abstract

Technology has been the catalyst that has facilitated an explosion of organisational data in terms of its velocity, variety, and volume, resulting in a greater depth and breadth of potentially valuable information, previously unutilised. The variety of data accessible to organisations extends beyond traditional structured data to now encompass previously unobtainable and difficult to analyse unstructured data. In addition to exploiting data, organisations are now facing an even greater challenge of assessing data quality and identifying the impacts of lack of quality. The aim of this research is to contribute to data quality literature, focusing on improving a current understanding of business-related Data Quality (DQ) issues facing organisations. This review builds on existing Information Systems literature, and proposes further research in this area. Our findings confirm that the current literature lags in recognising new types of data and imminent DQ impacts facing organisations in today's dynamic environment of the so-called "Big Data". Insights clearly identify the need for further research on DQ, in particular in relation to unstructured data. It also raises questions regarding new DQ impacts and implications for organisations, in their quest to leverage the variety of available data types to provide richer insights.

Keywords

Data Quality, Information Quality, Data Quality Issues, Unstructured Data, Literature Review

INTRODUCTION

Proliferation of use and penetration of technology in modern day society have invariably meant the variety, velocity and volume of data being created has been exponential (Eckerson, 2011). A dual effect of technological advances for the information organisation has been the ability to capture but also analyse a greater variety of data types. However, in order to benefit from a greater depth and breadth of potential information, its value is inherently dependent on the quality of the data being utilised.

Data and the information it provides to end-users could be considered the lifeblood of an organisation, influencing the overall effectiveness of organisational processes (Fox et al., 1999; McKendrick, 2011) and decision-making (Davenport et al., 2010). As organisations place greater reliance on deriving value from their information to be competitive (Lavalle, 2011; Eckerson, 2011), the impact of reduced DQ could be costly.

Various industry sources offer similar estimates that around 85% of an organisation's data, is unstructured (IBM as quoted by Abai, 2006; May, 2011). It is predicted that unstructued data will become more prevalent than traditional structured (relational) data (McKendrick, 2011). However, as a valuable information source unstructured data is currently under utilised in an organisational environment (McKendrick, 2011). While this issue requires organisational attention and focus, it is equally important to understand how users currently define data quality, especially new data types like unstructued data.

The ever expanding organisational data volumes are a result of the abundance of existing structured data and growth of unstructured data from "websites and machine generated data" including audio and video data (Eckerson, 2011). This growth in data variety has inevitably correlated to an increase in the range of applications and tools to utilise data. This in turn, has resulted in improved capabilities to integrate and analyse across a variety of data types.

However, to leverage the potential value of these different data types, it is imperative that organisations have an understanding of what constitutes quality data. Furthermore, failure to understand how a user perceives quality could result in costly implications for the organisation (Redman, 1998) especially as the use of different data types becomes more prevalent. "High quality data is critical to success in the Information Age", and lack of quality data could cost organisations 10-25% of total revenue (Eckerson, 2002). It's therefore important to consider data quality across all data types.

Focusing on the business, rather than technology perspective this research aims to explore the following research question:

What are the DQ issues in dynamic organisational environments?

This is further refined into the two more detailed research sub-questions

1) *How does an organisation understand and identify elements of quality data?* and

2) *What are the main impacts of DQ within organisations?*

Our analysis reveals that it is becoming increasingly important to extend the valuable contribution of prior research in the area of DQ. This is driven by an increased likelihood of new issues facing organisations due to changes in technology and the emergence of new data types from more sources and at a greater velocity, increasingly termed "Big Data" (May, 2011). Equally important is to ascertain if existing DQ frameworks are still applicable from an end users perspective or whether new characteristics are relevant and new impacts are now prevalent, calling for new theoretical frameworks.

The remainder of this paper is structured as follows. The following section introduces the key concepts used in this review. Next, the methodology used for this review is discussed. This is followed by the review and analysis of relevant literature. The final section summarises the main conclusions, states the limitations of this study and opens up a number of research questions for future research.

FOUNDATION CONCEPTS

Knowledge and understanding of data quality is critical, not only in isolation but also through the intrinsic relationship of data and quality. Numerous definitions of DQ and its characteristics can be found in current literature (Klein et. al., 1997; Fox et. al., 1994). However, before describing data as being of quality, it is first necessary to understand the "entity that's assumed to have a property called quality" (Lillrank, 2003), in this case data.

Considering the reliance and abundance of data available to organisations there is a lack of consensus on how to define data (Fox et. al., 1994). Although, characteristics of data have been described as "collected purposely, stored in a medium, repetitive in nature, and encoded in a specific format" (Nyaboga et al., 2009; supported by Lillrank, 2003; Levitin et al., 1995).

Organisational data types can be classified along a continuum ranging from structured to semi-structured to unstructured. However, for the purpose of this review the distinction between data types is provided Table 1.

Table 1: Classification of Data Type

Classification of Data Types		
Structured	**Unstructured**	
• Databases, Data Warehouses, Electronic Spreadsheets	• PowerPoint, Word Documents, Email, PDF, Audio files, Videos, Web Content, Graphics, Multi-media	• Web Posts, Blogs, Wiki pages, Forums, Tweets, Instant Messages

It's not disputed that semi-structured data has unique properties that warrant their inclusion in Table 1. However, adopting a bi-polar classification provides the necessary boundaries to confirm if quality characteristics change along this continuum.

Definitions of quality have been associated with products and people but could also encompass services and processes (van der Pijl, 1994). The determinants used to describe quality comprise a set of characteristics and in most cases considered in terms of its use and to satisfy a stated need (van der Pijl, 1994; Lillrank, 2003). Therefore data - the word preceding quality - is considered to be the change agent in this case.

Fusing both concepts - data and quality, a widely adopted view is that DQ is considered as being "fitness for use'" from the perspective of the end user (Lee, 2003; Ballou and Tayi, 1998; Wang et al., 1996) irrespective of data type. Due to the numerous characteristics proposed in the literature to describe data as being of quality, the concept of data quality is now is seen as multidimensional (Fox et al., 1994; Wang and Strong, 1996; Lee et. al., 1997; Levitin and Redman, 1995).

The organisational environment is now characterised by expanding data types, including traditional structured data and new types of unstructured data (Eckerson, 2011). For organisations both data types now provide an opportunity to add business value in isolation and in combination. Therefore, considering the intricate bond between data and quality, a re-assessment of our current understanding of data quality is required.

Finally, despite distinct meanings, some researchers such (Kokemuller, 2011) use the DQ and information quality (IQ) concepts interchangeably. Others such as (Price and Shanks, 2005) clearly distinguish between the two. In this project we primarily focus on data quality.

RESEARCH METHOD

This review analyses existing literature on DQ/IQ issues in organisations, taking a business rather than technology perspective. The research approach taken is an adaptation and combination of two methodologies, graphically depicted below in Figure 1. The foundation is provided by vom Bocke et al (2009) and supplemented with Webster and Watson (2002).

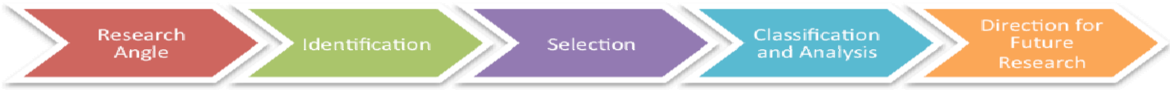

Figure 1: Research Method

1) Research angle (focus)

The Frame of Reference (Figure 2) identifies three research elements (a) topic, (b) scope and (c) perspective, with the intersection of these elements forming the research angle (focus). In our project, these elements included a) data quality as the main topic, b) organisational scope and c) business issues as our chosen perspective. The intersection of these elements defines our research focus: organisational DQ issues.

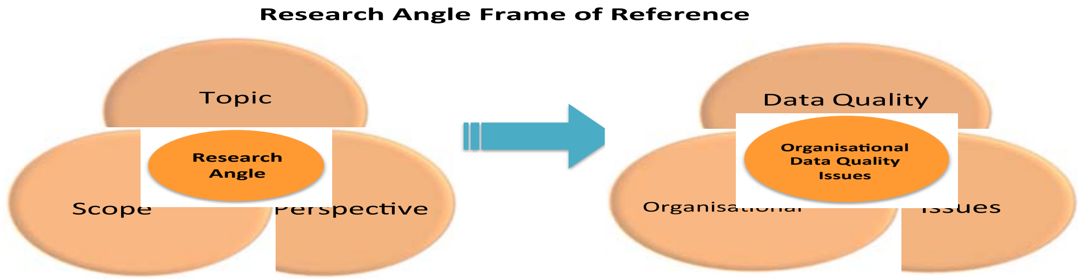

Figure 2: The Frame of Reference

2) Identification of information sources

Our literature review started from the top peer reviewed journals within the Information Systems domain - the so-called “senior scholars’ basket” of IS journals (AIS, 2009). This was supplemented with a niche journal to offer a broader view pertaining to the topic area and proceedings of renowned conferences (vom Brocke et al. 2009). The selected Association for Information Systems (AIS) sponsored and affiliated conferences ensured current perspectives and developments were captured. This included Americas Conference on Information Systems and International Conference on Information Systems, ranked the top two IS conferences by Walstrom and Hardgrave (2001).

The majority of papers (1085 papers out of 1573) identified were published in conference proceedings, “to some extent, this can be traced back to the time-consuming journal review process” (Wiener et al., 2010). Additional factors such as publication frequency and page capacity limitations may also contribute to this. Literature was constrained to the time frame defined by the utilised database. However, if deemed a foundation paper and published in the above sources, it was also included even if it was outside the specified time frame due to database constraints.

We readily acknowledge that a greater variety and number of alternative sources are available, such as other journals, books and conferences. However, the sources identified above were sufficient to gauge issues and themes pertaining to the research angle of this review.

3) Selection Process

All fields were searched using selected key words, “data quality” or “information quality” based on analysing current literature and the methodology of other research in this area such as Neely and Cook (2008). A total pool of 1537 papers, were initially identified. Due to the selected keywords being used interchangeable in literature meant that some papers appeared twice in the returned results, so they were eliminated from the pool.

Two selection filters were then employed. The initial filter omitted “conference papers that resulted in journal articles, papers with no original content such as announcements or forewords and research in progress papers” (Wiener et al., 2010), and editorials. The second filter involved skim reading of the title then the abstract of each paper to only include the papers with the main research focus on DQ/IQ. A forward and/or backward selection strategy was not used. This process reduced the number of relevant papers to 37. These resulting papers were then individually analysed, as follows.

Classification and Analysis Process

Literature classification was based on the Webster and Watson concept matrix approach (2002). Concept identification was achieved by skim reading current industry literature (including industry surveys and leading industry reports). It was also informed by the researchers’ professional knowledge and experience in the DQ area, with one of the co-authors being a practicing professional in this area. The identified concepts were organised into a tree like structure to observe their level of hierarchy and relationship.

Paper categorisation allowed the identification of themes, trends and current gaps. Although the “chosen approach adds a degree of subjectivity, it also adds richness from dealing with ideas and their recombination, rather than trying to solely infer meaning from quantitative attributes” (Wiener et al., 2010).

The research questions identified earlier formed the “analysis lens” through which the relevant literature was analysed and is reviewed in section 4 below.

4) Direction for Future Research

Following our in-depth analysis and synthesis of the relevant literature, we focused on identifying possible areas for future research, taking into account research gaps as well as current industry needs as identified in the recent industry reports, such as (Eckerson, 2011; Brown et al., 2011; McKendrick, 2011).

LITERATURE REVIEW

Understanding and Identifying Elements of Data Quality (Research sub-question 1)

The proliferation of technology and the ways organisations interact and utilise it has resulted in an exponential increase in organisational data, especially in recent times. While research has been undertaken to understand and define data quality, minimal research or discussion has sought to segregate data type in an organisational context as shown in Table 2.

Table 2: Research by Data Type

Data Type	Sources
All (assumed if unspecified)	(Moseley, 2010); (Rutstein, 2011); (Lee and Strong, 2003); (Madnick et al., 2003); (Cappiello et al., 2003); (Lee, 2003); (Wang and Strong, 1996); (Nelson et al., 2005); (Zhou et al., 1996); (Santos et al., 2010); (Kokemueller, 2011); (Jung et al., 2005); (Klein, 2003); (Klein, 2001); (Klein, 1999); (Becker et al., 2009); (Najjar and Bishu, 2005); (Nelson, 2002); (Kerr et. al., 2007); (Ge et. al., 2011); (Gorla et. al., 2010); (van der Pijl, 1994); (Price and Shanks, 2005); (Fisher et. al., 2003); (Ballou and Pazer, 1995); (Price and Shanks, 2011); (Klein et. al., 1997); (Wixom and Watson, 2001)
Unstructured	(Abai, 2006); (Klein, 1999); (Price and Shanks, 2005)

Various frameworks have been proposed to guide organisations understanding of data quality. However, judging by its publication date, the DQ mainstream literature was published well before industry-wide proliferation of new data types from new data sources. We envisage that this latency must raise questions and issues from an organisations perspective surrounding the relevancy and applicability of previously published DQ frameworks.

Furthermore, it is important to understand user perceived data quality in organisations that are currently characterised by increasing reliance and utilisation of structured and unstructured data to provide richer insights. Our analysis also reveal that organisations use two different approaches to assess the quality of their data - the end user or subjective approach, and system defined or objective approach (Ge et al. 2011). Key findings are shown in Table 3. Although bi-polar in nature, frameworks and perspectives occur along a continuum of these extremes to convey and discuss DQ characteristics.

Table 3: Research by Quality Perspective

Quality Perspective:	Sources:
User Defined (Subjective)	(Abai, 2006); (Dubois, 2005); (Lee and Strong, 2003); (Lee, 2003); (Wang and Strong, 1996); (Nelson et. al., 2005); (Klein, 1999); (Najjar and Bishu, 2005); (Fisher et. al., 2003)
User and System Defined (Objective/Subjective)	(Zhou et. al., 2006); (Santos, 2010); (Ge et. al., 2011); (van der Pijl, 1994); (Price and Shanks, 2005);
System Defined (Objective)	(Rutstein, 2011); (Madnick et. al., 2003);

Wang and Strong introduced a "hierarchical framework that captures the aspects of data quality that are important to data consumers" (1996). Taking a subjective approach they define DQ as being "fit for use" from the end users perspective (Wang and Strong 1996). Based on the end users perspective their framework comprised four categories. Intrinsic IQ focusing on data values, Contextual IQ in relation to the requirements for a specified task, Representational IQ focuses on its format and meaning and Accessibility IQ in terms of access and security of the data (Wang and Strong 1996).

A subjective approach is commonly used to define DQ determined by the user's perception of it being "fit for use" (Lee et al. 2003; Wang et al. 1996). Implications of this approach from an organisational perspective are threefold. Firstly, ensuring the completeness and correctness of characteristics (Wang et. al., 1996). Secondly, understanding the judgements made about quality from differing parties. Thirdly, organisations need to understand the "difficulty in reliably measuring and quantifying such perceptions" (Price and Shanks, 2005).

Although this approach inherently implies a level of interpretation by the user it does provide organisations a greater richness in assessing quality and revealing unconsidered characteristics (Wang et al., 1996). This is an important consideration in an organisational environment now characterised by an evolving use of multiple data types including unstructured data.

Other frameworks integrate an objective and subjective approach to define data quality. Price and Shanks (2005) use semiotic theory to propose the so-called semiotic information quality framework. The framework comprised three levels; syntactic, semantic, and pragmatic from the viewpoint of quality categories and characteristics. The syntactic level deals with the conformance to database rules. The semantic level deals with the information related to external phenomena while the pragmatic level focuses on information use (Price and Shanks, 2005). Ge et al. proposed a multi-dimensional framework involving three layers: evaluators, assessment dimensions and assessment target (2011). They defined evaluators as being users or software, assessment dimensions were categorised as Acquisition IQ, Context IQ, Specification IQ and Expectation IQ and assessment target depended if it is an information product or raw data (Ge et al., 2011).

It could be debated that it is first necessary to understand the viability to improve the quality of all data types including unstructured data. But at the same time, it might be necessary to understand how users define the quality of all data types first, before one could determine the feasibility of any data quality improvement effort.

Irrespective of the approach taken, our research confirms that when it comes to framework development, data type is seldom mentioned. We also found out that the existing frameworks and research describing DQ use a number of common characteristics such as accuracy, consistency, timeliness, completeness and accessibility (Zhou et al. 2006; Lee and Strong 2003, Cappiello et al. 2003; Wang and Strong. 1996). The importance of these characteristics are not challenged, however, understanding of users perceptions is only relative to the accessibility of data types at that point in time. As information value is now derived from structured and unstructured data, it is important to re-examine our current knowledge of user perceived data quality by data type.

Furthermore, most of the existing frameworks found in the literature describe and classify the characteristics of DQ. Some go a step further and consider mutual dependencies among DQ characteristics. For example, in the context of decision-making the Ballou and Pazer framework discussed that optimal DQ is achieved by users via trade-offs between accuracy and timeliness (1995). Whereas Cappiello et. al. (2003) discuss characteristic trade-offs in multi-channel information systems between time related accuracy and time related completeness.

Given the increasing pressure to utilise data efficiently, organisations need to understand these characteristic trade-offs and their impact. From an organisational perspective it is necessary to consider the trade-off between the cost of achieving quality and the output value of that quality. This inherently impacts end users and the trade-offs they make as the pressure increases to reduce decision latency. As data types become more integrated and utilised within an organisational environment, it is possible that new DQ characteristic trade-offs are likely to emerge.

Over time technology has shifted focus from traditional 'bricks and mortar' business models to facilitate the emergence of new business models. The presence of global economy and collaborative technology, places emphasis on exploring data quality characteristics as a result of these new inter-organisational partnerships. In response to economic globalisation Zhou et al (2006) developed a framework to guide organisational understanding in light of a changing organisational environment. The framework was developed along five categories, data integrity, data integration, data synchronisation, data transparency and data dynamics (Zhou et al., 2006).

Technology has allowed the leveraging of multiple data types to provide richer insights through integration and organisational connectivity. A very recent global survey of nearly 3000 executive managers and analysts, conducted by MIT Sloan and IBM (LaValle et al., 2011) confirms that data integration across functional and organisational boundaries remain the highest data-related priority. "One big challenge is the fact that the mountains of data many companies are amassing often lurk in departmental "silos", such as R&D, engineering, manufacturing, or service operations – impeding timely exploitation." (Brown et al., 2011). High quality of isolated (e.g. departmental) data, does not guarantee high quality of integrated data. However, current literature does not focus on this aspect.

The inherent unique properties of differing data types now possible are again expected to raise questions surrounding the applicability and relevancy of frameworks and quality characteristics. As organisations progress in the utilisation of data types (structured and unstructured) to provide value, hearsay provides a weak foundation to infer the applicability of these frameworks in today's organisational environment.

Impacts of Data Quality (Research sub-question 2)

The relationship between data quality and organisational impacts is typically discussed in the context of Information Systems (IS) success. Whereby data quality is discussed in the form of information quality and is one of several influencing factors including system quality (Wang and Strong, 1996; Ditzel et al., 2010; Gorla et al., 2010; Kokemuller 2011; Wixom and Watson, 2001; Jung et al., 2005) and service quality.

Impacts can arise either independently or in combination with other constructs such as system and service quality (2010). Extending data beyond their lowest form of words and numbers to incorporate the insights and actions of social actors (i.e. data consumers), is likely to create even greater effect. These effects can be discussed in terms of organisational impacts or individual impacts (Ditzel et al., 2010). Despite different perspectives correlation exists in some identified impacts, decision-making, productivity, innovation, customer satisfaction and management control (Ditzel et al., 2010).

Research into the impact between data quality and its effect on organisations is still emerging (Ditzel et al., 2010). It is suggested that unstructured data represents about 85% of organisational data (IBM cited in Abai 2006; May 2010) therefore by default 15% is structured data. Despite a shift in the growth and use of a variety of organisational data, impact statistics and issues tend to utilise research conducted in the 1990's and early 2000's. As a result it is inferred that these impacts pertain to only 15% of organisational data.

This also means that our current knowledge of organisational impacts due to lack of DQ are understated and/or present a narrow view. Latency in this area of research it's inferred is attributed to difficulties in separation of probable causes and quantification. This inevitably will be further complicated as the variety of data types being utilised increases within and across organisational boundaries. The unique properties of differing data types now present new challenges for organisations trying to gain a complete understanding of the impacts of data quality.

It's reasonable to assume technology has facilitated the growth in data variety. This in turn leads to some important questions, as follows. Has technology acted as a catalyst or minimised the impacts of data quality across data types? Alternatively, has technology had a dual effect, increasing available data types, while reducing the impact of poor data quality across data types? Through technology are new impacts arising for organisations from the growth, variety and speed at which data is being created, captured and stored? Is information hoarding emerging as a phenomenon, creating new impacts? If organisations currently fail to understand end users perceptions of data quality (by type) as it arises, and what are the likely consequences?

Organisations are increasingly facing pressure to deliver "business decisions, with less risk, while delivering better results" (Moseley, 2010). However, a common organisational impact of data quality in literature pertains to decision-making making (Abai, 2006; Kerr, 2007; Nelson, 2002; Jung et al., 2005; Becker et al., 2009; Kerr et al., 2007; Price and Shanks, 2005). How data quality impacts decision-making is discussed by Ballou and Pazer (1995) in terms of the trade-offs of two data quality characteristics; accuracy and timeliness. However, organisations still need to consider that its data as precursor to information quality is just one of several factors resulting in organisational impacts (noted above), because even high quality data could produce poor decisions.

Arguably, a key enabler to more informed organisational decision-making is via greater utilisation and leveraging of more organisational data types, beyond highly structured data. But at the same time, in dynamic environments with increased data variety and connectivity within and across organisations, the impact of data quality on decision performance is likely to increase.

A possible solution offered in the literature is to allow end users to assess data quality (Rutstein, 2011; Moseley, 2010; Kerr, 2007; Klein, 1997; Price and Shanks, 2011; Fisher et al. 2003), through data quality tagging. For example, Fisher et al. (2003) looked at the impact of experience and time pressures on decision makers and the usefulness of data quality tagging.

However, while offering the end users an opportunity to make their own assessment, the same solution potentially creates a myriad of previously unconsidered characteristic trade-offs. Debatably, it would be an inhibitor, adding additional time to the analysis process and increasing information and cognitive load to compound task complexity (Kerr, 2007). Further its representation will impact decision-making, as Jung et. al. (2005) indicated, there is a strong correlation between the effects of representational DQ and task complexity on decision performance. Additionally, the same solution would also be subject to quality assessment, creating potentially unidentified quality characteristics and impacts.

As the ‘lifeblood of an organisation’, the impact of data quality on business performance is another common topic. Focus is placed internally within and across functional barriers (Cappiello et al., 2003), from a traditional structured data perspective. However in a global market place the impact of data quality has now extend to inter-organisational business processes (Zhou et al., 2006; Najjar and Bishu, 2005). This extension and cross-utilisation of data through inter-organisational processes is expected to raise new questions surrounding impact and quality characteristics. Collaboration and data sharing via inter-organisation information systems, is resulting in new data quality issues such as inter-organisational data standardization and synchronization (Zhou et. al., 2006). But to what extent does data quality impact some internal business processes especially customer facing processes? This important question remains unexplored in the current literature.

CONCLUSIONS, LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Our literature review confirmed a research gap caused by the misalignment between current technology capabilities and organisational needs on one side, and the relevant DQ research published by the scientific community. However, even though most DQ papers were published well before wide proliferation of new types of data, many of the research questions identified previously remain unanswered and therefore still valid, if not even more relevant. In particular the role DQ plays as a key factor in the optimisation, utilization and exploitation of organisational data, including the resultant impacts of poor data quality.

From an organisational perspective of paramount importance will be further research to understand some of the below questions:

- Are current data quality frameworks still applicable to new data types?
- Is there alignment of quality characteristics between data types taken in isolation and in integrated forms?
- What are the organisational impacts of unstructured data and cross utilisation of data types?

Although, recognizing the impact of DQ is important for organisations, the pre cursor to this however, is to understand and identify the characteristics of quality from the end user. This is important given the already wide proliferation of business analytics tools aimed at end users, especially customer-facing employees.

The purpose of this review is not to de-value or dispute the importance of the existing body of knowledge on DQ but to encourage a new perspective and offer direction for further research relating to DQ issues facing organisations today. In spite of a very large pool of papers being considered in this research, we acknowledge that DQ issues are also considered by other research communities beyond information systems. There lies the main limitation of this research that we readily acknowledge.

We envisage that the insights offered here could have significant implications for organisations as our research confirms the need for further research on DQ, especially with regards to new data types. This research also raises questions surrounding the relevancy and applicability of existing frameworks and characteristics, prevalence of unidentified impacts.

Although a glacial shift in research focus is yet to occur from quality of structured data to unstructured data (Madnick, citied in Blake 2010), the importance of understanding DQ from an organisational perspective is however gaining momentum (Neely and Cook, 2008). This serves to not only frame the importance of the above insights but act as a catalyst to provide organisations with greater insights to help guide them (Neely and Cook, 2008) through the DQ “mine” field.

Our current and future research include re-visiting some of the well-established DQ/IQ frameworks and testing their validity in the world of new data types and new, more complex requirements for data quality in dynamic organisations, especially in relation to "Big Data" as defined by (May, 2011). Given the fact that new technologies such as real-time business analytics are increasingly used to support decision-making processes by end-users (e.g. customer-facing employees) rather than by a group of dedicated expert business analysis, we are particularly interested to contribute to better understanding of end-user perception of data quality and their organisational implications.

REFERENCES

Abai, M. 2006. "Building a Data-Centric Organisation", *Business Intelligence Journal* (11:4), Fourth Quarter, pp 30-36.

Ballou, D., and Pazer, H. 1995. "Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff", *Information Systems Research* (6:1), March, pp 51-72.

Ballou, D., and Tayi, G. 1998. "Examining data quality", *Communication of the ACM* (42:2), February, pp 54+.

Becker, J., and Poeppelbuss, J., and Gloerfeld, F., and Bruhns, P. 2009. "The Impact of Data Quality on Value Based Management of Financial Institutions", *In Proceedings of Americas Conference on Information Systems (AMCIS)*, P.490, California, San Francisco.

Blake, R. 2010. "Identifying the Core Topics and Themes of Data and Information Quality Research", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P221, Peru, Lima.

Blumberg, R., and Atre, S. 2003. "The Problem with Unstructured Data", *Information Management*, Retrieved 20 July 2012 from www.information-management.com.

Brown, B., and Chui, M., and Manyika, J. 2011. "Are you ready for the era of "big data"? *McKinsey Quarterly*, October 2011.

Cappiello, C., and Fancalanci, C., and Pernici, B. 2003. "Time-Related Factors of Data Quality in Multichannel Information Systems", *Journal of Management Information Systems* (20:3), Winter, pp 71-91.

Davenport, T.H., and Harris, J.G. and Morison, R. 2010. *Analytics at Work*, Harvard Business Review Press.

Ditzel Santos, G., and Takaoka, H., and de Souza, C. 2010. "An Empirical Investigation of the Relationship between Information Quality and Individual Impact in Organisations", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.192, Peru, Lima.

Eckerson, W. 2002. *Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data*, Research Report, The Data Warehousing Institute.

Eckerson, W. 2011. *Big Data Analytics: Profiling the Use of Analytical Platforms in User Organisations*, BeyeNetwork.

Eppler, M., and Wittig, D. 2000. "Conceptualising Information Quality: A Review of Information Quality Frameworks form the Last Ten Year", *In Proceedings of the 2000 Conference on Information Quality*, USA, Massachusetts, pp 83-96.

Fisher, C., and Chengalur-Smith, I., and Ballou, D., 2003. "The Impact of Experience and Time on the Use of Data Quality Information in Decision Making", *Information Systems Research* (14:2), June, pp 170-188.

Fox, T., and Guynes, C., and Prybutok, V., and Windsor, J. 1999. "Maintaining Quality in Information Systems", *The Journal of Computer Information Systems* (40:1), Fall, pp 76-80.

Fox, C., and Levitin, A., and Redman, T. 1994. "The Notion of Data and its Quality Dimensions", *Information Processing & Management* (30:1), January-February, pp 9-19.

Ge, M., and Helfert, M., and Jannach, D. 2011. "Information Quality Assessment: Validating Measurement Dimensions and Processes", *In Proceedings of European Conference on Information Systems (ECIS),* P.75, Finland, Helsinki.

Gorla, N., and Somers, T., and Wong, B. 2010. "Organisational Impact of System Quality, Information Quality, and Service Quality", *The Journal of Strategic Information Systems* (19:3), September, pp 207-228.

Jung, W., and Olfman, L., and Ryan, T., and Park, Y. 2005. "An Experimental Study of the Effects of Representational Data Quality on Decision Performance", *In Proceedings of Americans Conference on Information Systems (AMCIS),* P.298, Nebraska, Omaha.

Kerr, K., and Norris, T., and Stockdale, R. 2007. "Data Quality Information and Decision Making: A Healthcare Case Study", *In Proceedings of Australasian Conference on Information Systems (ACIS),* P.98, Toowoomba.

Klein, B. 1999. "Perceptions of Information Quality: A study of Internet and Traditional Text Sources", *In Proceedings of Americas Conference on Information Systems (AMCIS)*, P.214, Wisconsin, Milwaukee.

Klein, B. 2001. "Incentives, Base Rate Expectations, and False Alarms in the Detection of Data Errors", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.74, Massachusetts, Boston.

Klein, B. 2001. "User perceptions of data quality: Internet and traditional text sources", *The Journal of Computer Information Systems* (44:4), Summer, pp 9-15.

Klein, B. 2003. "Some Determinants of User Perceptions of Information Quality on the World Wide Web", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.295, Florida, Tampa.

Klein, B., and Goodue, D., and Davis, G. 1997. "Can Humans Detect Errors in Data? Impact of Base Rates, Incentives, and Goals", *MIS Quarterly* (21:2), June, pp 169-194.

Kokemueller, J. 2011. "An Empirical Investigation of Factors Influencing Data Quality Improvement Success", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.154, Michigan, Detroit.

LaValle, S., and Lesser, E., and Shockley, R., and Hopkins, M., and Kruschwitz, N. 2011. "Big Data, Analytics and the Path from Insights to Value", *MITSloan Management Review* (52:2), Winter, pp 21-31.

Lee,Y., and Strong, D., and Wang, R. 1997. "Data quality in context", *Communication of the ACM* (40:5), May, pp 103+.

Lee, Y., and Strong, D. 2003. "Knowing-Why about Data Processes and Data Quality", *Journal of Management Information Systems* (20:3), Winter, pp 13-39.

Lee, Y. 2003. "Crafting Rules: Context-Reflective Data Quality Problem Solving", *Journal of Management Information Systems* (20:3), Winter, pp 93-119.

Levitin, A., and Redman, T. 1995. "Quality Dimensions of a Conceptual View", *Information Processing & Management* (31:1), January-February, pp 81-88.

Lilllrank, P. 2003. "The Quality of Information", *The International Journal of Quality & Reliability Management* (20:6), pp 691-703.

Madnick, S., and Wang, R., and Xian, X. 2003. "The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Quality", *Journal of Management Information Systems* (20:3), Winter, pp 41-69.

May, T. 2011"Big Data/Big Analytics", retrieved on 20th July 2012 from http://www.youtube.com/watch?v=GmR7BM5B8Wk

McKendrick, J. 2011. "The Post-Relational Reality Sets in: 2011 Survey on Unstructured Data", Unisphere Research.

Moseley, M. 2010. "Improving Data Quality through Agile Data Governance", *Business Intelligence Journal* (15:1), First Quarter, pp 17-24.

Najjar, L., and Bishu, R. 2005. "Information Quality in the Banking Industries", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.275, Nebraska, Omaha.

Neely, M., and Cook, J. 2008. "A Framework for Classification of the Data and Information Quality Literature and Preliminart Results (1996-2007)", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.131, Canada, Toronto.

Nelson, K. 2002. "Determinants of Data Quality Management Adoption and Implementation", *In Proceedings of Americas Conference on Information Systems (AMCIS),* P.20, Texas, Dallas.

Nelson, R., and Todd, P., and Wixom, B. 2005. "Antecedents of Information and Systems Quality: An Empirical Examination within the Context of Data Warehousing", *Journal of Management Information Systems* (20:4), Spring, pp 199-235.

Nyaboga, A., and Mwaura, M. 2009. "Strategies For Gaining Competitive Advantage In A Dynamic Environment Thru Data Quality", *International Journal of Management and Information Systems* (13:1), April, pp 13-21.

Price, R., and Shanks, G. 2005. "A Semiotic Information Quality Framework: Development and Comparative Analysis", *Journal of Information Technology* (20:2), June, pp 88-120.

Price, R., and Shanks, G. 2011. "The Impact of Data Quality Tags on Decision-Making Outcomes and Process", *Journal of the Association for Information Systems* (12:4), April, pp 323-346.

Redman, T. 1998. "The Impact of poor data quality on the typical enterprise", *Communications of the ACM* (41:2), February, pp 79-82.

Roberts, P. 2010. "*Corralling Unstructured Data for Data Warehouses*", *Business Intelligence Journal* (15:4), pp 50-55.

Rutstein, J. 2011. "Is it Really "Bad" Data?", *Business Intelligence Journal* (16:3), September, pp 7-13.

Walstrom, K., and Hardgrave, B. 2001. "Forums for Information Systems Scholars: III", *Information & Management* (39:2), December, pp 117-124.

Wang, R., and Strong, D. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers", *Journal of Management Information Systems* (12:4), March, pp5-34.

Webster, J., and Watson, R. 2002. "Analysing the Past to Prepare for the Future: Writing Literature Review", *MIS Quarterly* (26:2), June, xiii-xxiii.

Weiner, M., and Vogel, B., and Amber, M. 2010. "Information Systems Offshoring-A Literature Review and Analysis", *Communications of the Association for Information Systems* (27:1), pp 455-492.

Wixom, B., and Watson, H. 2001. "An Empirical Investigation of the Factors Affecting Data Warehousing Success", *MIS Quarterly* (25:1), March, pp 17-41.

Van der Pijl, G J. 1994, "Measuring the strategic dimensions of the quality of information", *Journal of Strategic Information Systems* (3:3), September, pp 179-190.

Vom Brocke, J., and Simons, A., and Niehaves, B., and Riemer, K., and Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process", *In Proceedings of European Conference on Information Systems (ECIS)*, P.161, Italy, Verona.

Zhou, D., and Chuang, T., and Nakatani, K. 2006. "A Conceptual Data Quality Framework for Collaborative Commerce", *In Proceedings of Americas Conference on Information Systems (AMCIS)*, P. 56, Mexico, Acapulco.

COPYRIGHT