The Value of Good Data - A Quality Perspective
A Framework for Discussion

Tony O'Brien[1], Arun Sukumar[1] and Markus Helfert[2]
[1]*Sheffield Business School, Sheffield Hallam University, Howard Street, Sheffield, U.K.*
[2]*Department of Computing, Dublin City University, Dublin, Ireland*

Keywords: Data Quality, IT Governance, Data Assurance, IT Management, Cost Benefits, Information Value Management.

Abstract: This study has highlighted the benefits and value of quality information and the direct consequences associated with low quality data. This paper also describes a number of taxonomies which may be used to classify costs relating to both the consequences of low quality data and the costs of improving and assuring on-going data quality. The study then provides practical examples of data quality improvement initiatives undertaken within two large organisations. Finally a data governance model is proposed centring on three inter-related fundamental elements namely: People, Processes and Data, where any attempt to improve the quality of data within any organisation must be focussed around these three essential elements.

1 INTRODUCTION

Many researchers within the context of management planning and information systems have identified the importance of data (Davenport, 1998; Davenport et al., 2001; Galliers and Newall, 2001; Davenport and Harris 2002; Newall, Huang, Galliers and Pan 2002; Davenport, Harris and Cantrell 2004). From this research a realisation has grown that organisations that are able to collect, analyse and act on data in a strategic manner, are in a position to gain a competitive advantage within their industries, leading in some cases to domination in these areas (Davenport, 2006). This form of information management known as 'analytics' stresses that successful organisations are those that take action from their information to inform their strategic decision making Davenport (1998); Davenport et al., (2001); Davenport (2006); Davenport and Harris (2007) Davenport (2009), establishing along the way a 'fact-based culture' (Harris 2005a; Harris 2005b; Harris 2007).

If this ever expanding focus on 'intelligent' business intelligence and management information is so crucial to organisational strategy, then the requirement to have quality data becomes even more paramount in manufacturing planning (Gustavsson and Wanstrom (2009: 326) as well as information systems (Davenport, Harris and Cantrell 2004: 23; Stenmark 2004: 1; Economist Intelligence Unit 2006: 2, 16; Foley and Helfert 2010: 477; Davenport, Harris and Morison 2010: 1). Over the last two decades data quality has been identified as a major concern for many enterprises Redman (1995); English (1998); Redman (1998); English (1999); Loshin (2001); Redman (2001); Eckerson (2002); Redman (2002); Redman (2004); English (2009), none more so than those operating enterprise resource planning and information systems (Deloitte 1999).

A report from The Data Warehouse Institute estimated that data quality problems costs US business $600 billion a year (5% of the American GDP) in postage, printing and staff overhead costs alone, whilst the majority of the senior managers in those companies affected remained unaware (Eckerson, 2002: 3). More recently English (2009: 4-15) outlined a catalogue of corporate disasters emanating from poor quality business information amounting to 'One and a Quarter Trillion Dollars' (English 2009: 15). During 2009 a survey of 193 organisations sponsored by Pitney Bowes, 39% of which had revenues in excess of US $1 billion, reported that a third of the respondents rated their data quality as poor at best, whilst only 4% reported it as excellent (Information Difference, 2009: 4).

A further survey found that less than one third of organisations regularly monitor data quality (Hayter

O'Brien T., Sukumar A. and Helfert M..
The Value of Good Data - A Quality Perspective - A Framework for Discussion.
DOI: 10.5220/0004616805550562
In *Proceedings of the 15th International Conference on Enterprise Information Systems* (IVM-2013), pages 555-562
ISBN: 978-989-8565-60-0

2010: 22), whilst a Gartner report stated that "Through 2011, 75% of organisations will experience significantly reduced revenue growth potential and increased costs due to the failure to introduce data quality assurance" (Fisher 2009: 6).

Strong parallels have been within the research literature between the concept of a planning and information system and that of a manufacturing system (Strong et al., 1997:104; Wang, 1998: 59 and latterly Pham Thi and Helfert, 2007: 6). The principle elements are highlighted below within what may be termed an Information Process Model see Table 1 to compare and contrast the various elements:

The 'Manufacturing' or 'Factory' analogy is a useful model in that it takes a conceptual over-view of both generic manufacturing and information systems to identify ways in which established quality principles may be applied to the input and process elements ensuring that information products in the form of outputs conform to the requirements of their relevant customers. Strong, Lee and Wang (1997:104) identified three key roles within a data manufacturing system:

Data Producers: Generate data
Data Custodians: Manage, store and process data
Data Consumers Use data and information

Within this context, however, one needs to be aware that the end products from manufacturing and information systems have differing implications, with the information production process viewed as potentially a more complex process than its physical equivalent (Pham Thi and Helfert 2007: 6). The outputs from a factory are unique one-off products which can be consumed only once, whether they are finished goods or components requiring further work. The overall effects of poor manufacturing are somewhat limited, normally requiring a scrap and re-work operation. Some longer-term detrimental implications may occur including customer dissatisfaction or product contamination, but even these will normally be relatively localised and time-constrained. Output in the form of data or information products can be consumed in an infinite number of ways and be re-cycled continually. Poor data can act like a virus infiltrating all aspects of an enterprise's operations, re-occurring again and again, or lay hidden undetected within sub-systems in perpetuity. Data may also be used in ways for which it was not created or intended, causing potential misalignment, errors or misinterpretations, resulting in potentially dangerous or catastrophic decision making (Senge, 1992: 7; Orna, 2005: 44, 144-150; Mutch, 2008: 53).

2 RESEARCH APPROACH

Our research process to develop the proposed framework can be seen as elements of Design Science-oriented research process (Braun et al. 2005; Hevner et al. 2004). In this paper we scoped the problem and based our findings on reviewing relevant literature and feedback from case studies. The literature was collected primary from journals and prominent book contributions related to design science. The literature review was complemented by a series of discussion-type focus group meetings with domain experts sharing experiences and challenges. This approach attempted to generate discussion and interaction to confirm our framework. The use of a design science oriented research approach in this environment provided the study with a considerable degree of richness. From the outset certain important notions and impressions emerged from the discussions and the analysis and these were subsequently developed as key findings. In the following we present first findings from our literature review and conceptual framework.

3 BENEFITS AND VALUE OF DATA QUALITY

The second element to be considered when evaluating data quality is its impact and value to the business. The aspect of business value in relation to IS has been discussed in numerous papers. For instance, Gustafsson et al. (2009) have presented a comprehensive model that aims to explain the business impact with three generic elements: IT, organizational impact, and business value. This model serves as background model for data quality impact. Other related frameworks have been presented in the literature aiming to refine this generic model (Borek et. al., 2011).

The model is supported by strong evidence that data quality has a considerable effect on decision-making in organizations. This section will therefore focus on the data quality value in decision-making. For instance, Keller and Staelin (1987) indicate that increasing information quantity impairs decision effectiveness and, in contrast, increasing data quality improves decision effectiveness. Jung et al. conducted a study to explore the impact of

Table 1: Information Processing Model.

Generic Process	Manufacturing System	Generic Information System	ERP Environment
Input	Raw materials	Raw data	People- Processes-Data
Process/ Operations	Production line	Information system	ERP Database
Output	Physical products	Information products	Information-People

representational data quality (which comprises the data quality dimensions interpretability, easy to understand, and concise and consistent representation) on decision effectiveness in a laboratory experiment with two tasks with different levels of complexity (Jung and Olfman, 2005). Furthermore, Ge and Helfert (2008) show that the improvement of data quality in the intrinsic category (e.g. accuracy) and the contextual category (e.g. completeness) can enhance decision quality.

Further studies elaborate this theme. English (1999) divides data quality costs into three main characteristics, costs caused by low data quality, assessment and inspection costs incurred to verify if processes are performing properly and process improvement and prevention costs. Loshin (2001) focusses upon the effects of low quality data in terms of its impact over time in relation to the traditional levels of organisational decision making. Shorter term operational impacts covering aspects of a system for processing information together with the costs of maintaining such operational systems involving elements of detection, correction, reworks and ultimate prevention. Medium term tactical aspects which attempt to anticipate issues and problems and finally long term planning where the impact of poor data quality can delay important strategic decisions resulting in lost opportunities and overall poor strategic decision making (Loshin 2001).

Haug et al., (2011) followed up the research of Ge and Helfert (2007), who identified three major components relating to this area: (1) information quality assessment, (2) information quality management, and (3) contextual information quality as shown in Table 2.

Following on from the themes of English (1999) and Loshin (2001) the Eppler and Helfert model dissects data quality costs into two major classifications relating to those costs incurred as a result of low quality data and the consequential costs of improving or assuring ongoing data quality. Each classification then consists of subordinate categories relating to the direct and indirect costs of poor data and the prevention, detention and repair costs associated with data quality improvement processes as shown in Table 3.

Each subordinate category is then further subdivided into six quality costs element and seven cost improvement elements. In order to investigate the business value of data quality, we follow IS/IT business value studies that show how IS/IT impacts on business processes and/or decision-making. A business process can be defined “a specific ordering of work activities across time and place, with a beginning, an end, and clearly identified inputs and outputs: a structure for action” (Davenport, 1993). Porter and Millar argue that activities that create value consist of a physical and an information-processing component and each value activity uses information (Porter and Millar, 1985).

In their integrative model of IS/IT business value, Mooney et al. (1996) propose a process framework for assessing the IS/IT business value. They present a typology of processes that subdivides business processes into operational and management processes and argue that IS/IT creates business value as it has automational, informational, and transformational effects on the processes. Similarly, Melville et al. (2004) see business processes and business process performance as the key steps that link organizational resources to organizational performance. Data can be seen as an important organizational asset as well as resource. Its quality is directly related to business value and organizational performance. In addition to measuring the effect on business processes, organizational performance has always been of consideration to IS/IT researchers and practitioners, resulting in a plethora of performance related contributions. Earlier approaches focused, for example, on the economic value of information systems (Van Wegen and De Hoog, 1996). They were more recently detailed to frameworks for assigning the impact of IS/IT to businesses (Mooney et al., 1996; Melville et al., 2004). These IS/IT oriented frameworks have resulted in an abundance of recommendations, frameworks and approaches for performance measurement systems (Folan and Browne, 2005).

Table 2: "Classification of information quality problems identified in literature". Source: Ge and Helfert (2007). Adapted by (Haug et al., 2011).

	Data Perspective	User Perspective
Context-independent	Spelling error Missing data Duplicate data Incorrect value Inconsistent data format Outdated data Incomplete data format Syntax violation Unique value violation Violation of integrity constraints Text formatting	The information is inaccessible The information is insecure The information is hardly retrievable The information is difficult to aggregate Errors in the information transformation
Context-dependent	Violation of domain constraint Violation of organization's business rules Violation of company and government regulations Violation of constraints provided by the database administrator	The information is not based on fact The information is of doubtful credibility The information presents an impartial view The information is irrelevant to the work The information consists of inconsistent meanings The information is incomplete The information is compactly represented The information is hard to manipulate The information is hard to understand

Table 3: "A data quality cost taxonomy". Source: Eppler and Helfert (2004). Adapted by (Haug et al., 2011).

Data quality costs	Costs caused by low data quality	Direct costs	Verification costs
			Re-entry costs
			Compensation costs
		Indirect costs	Costs based on lower reputation
			Costs based on wrong decisions or actions
			Sunk investment costs
	Costs of improving or assuring data quality	Prevention costs	Training costs
			Monitoring costs
			Standard development and deployment costs
		Detection costs	Analysis costs
			Reporting costs
		Repair costs	Repair planning costs
			Repair implementation costs

It has been recognized that there are two perspectives on value: objective and perceived value, which results in different data quality and value measures and value perceptions for particular stakeholders (Fehrenbacher and Helfert, 2012). To evaluate the value of data quality and develop suitable indicators, we suggest combining the work on business processes and decision quality with the work on performance indicators, developing a framework for analyzing business value of data quality. This is illustrated in Figure 1. The value propositions of data quality are manifold. It ranges from generating direct business value by providing information of high quality, reducing complexity, improving customer loyalty, improving operational performance, reducing costs and so on. Due to its critical importance in enterprises, data quality affects many areas of organizational performance and may deliver business value simultaneously to stakeholders.

4 CASE EXAMPLES

This article has provided illustrations from the literature to highlight examples of the costs of poor data quality and consequential benefits of related improvement programmes. A further example of the effects of such an initiative may be seen from a practical data quality improvement programme allied to an academic study carried out in collaboration with an industry partner. This organisation, a multi-business manufacturing enterprise operating across sixty three factories and offices within the United Kingdom, initiated a data quality improvement programme in 2006 and over the subsequent five years the quality of its overall data as measured by a weighted KPI index showed an overall improvement of 59%.

Looking at the cost taxonomy, whilst there was no detailed analysis undertaken as to the detailed financial effects of the underlying data quality problems, the improvement initiative was undertaken by existing staff using existing resources, applying quality principles which evolved during the overall process. These were basically 'sunk costs' in that there were little or no marginal incremental costs incurred as a direct consequence of the overall initiative. Whilst it could be argued that such resources could have been applied in other areas of the business, the overall effects upon the business mean that data was identified as a major organisational resource and asset.

During the period the overall operating results improved by 37%, with a 52% improvement in operational order efficiency across purchasing, manufacturing and sales/despatches. In addition the underlying problems in processing supplier invoices and successfully resolving customer invoice issues improved by 72% and 53% respectively.

Whilst the links between the data quality initiative and improved financial position could be somewhat tenuous, it was widely acknowledged within the organisation that the operational improvements were a direct result of the programme.

A similar study conducted more recently on a large quasi-public sector organisation has again highlighted the costs of poor data quality. The organisation, used to be one of the largest public sector organisation has recently been privatised and has faced numerous problems relating to data quality whilst providing its services. The study conducted in the form of focus groups, highlighted a number of key themes relating to data quality. The main themes identified are given as follows,

Firstly, in the discussion among the cross section of the work force, it was noted that data and information governance were of low priority. Employees' awareness of data governance issues and the associated responsibilities were low; the communication channels that are used to highlight

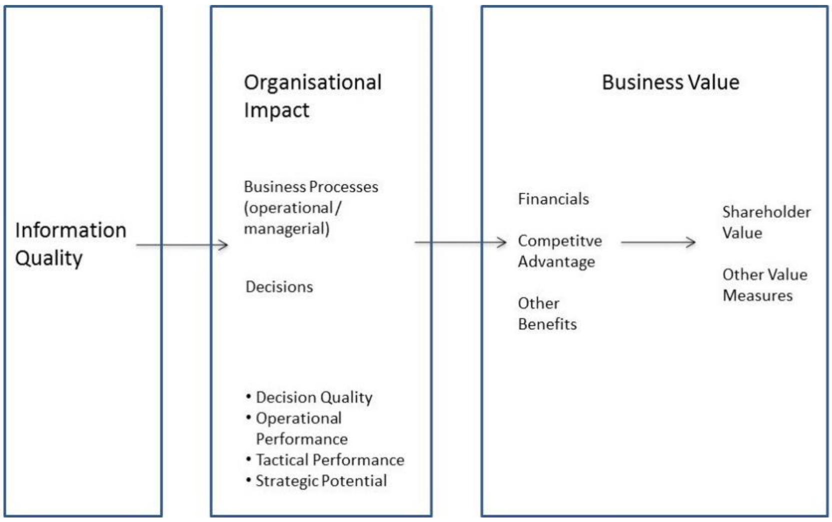

Figure 1: Framework for analyzing business value of data quality.

and promote data quality issues are either non-existent or clogged. Secondly, there was an absence of any formal mechanism or a procedure to report data problems. Employees who worked with the Master Data Systems were not aware of any formal procedures or mechanisms through which they can report or correct faulty or incorrect data. One attendant stated that "When I send payments to subxxxxx, I am not even sure that the branch is still open, I may actually be sending payments to the wrong person or to a wrong branch". Thirdly, the organisation did not have formal structure in terms of data stewardship or governance, data management was done on an ad-hoc basis by senior managers and specific roles and responsibilities relating to data quality management were either absent or under-developed. Fourthly, a more common theme identified related to the use of local and informal controls to manage data. Examples included the use of local spreadsheets, storing mission critical data in local drives, users writing their own macros to automate some actions etc. These issues though provide convenience and expedite the transactions in local areas can often lead to information security risks and compliance issues. Lastly, among the discussions it was noted that the middle level managers were not aware of ISO standards or best practices associated with information security management, They were aware of the need to employ and use the current best practices available within the information security management domain but the knowledge to get further relevant information or how to implement a organisation wide data management program was lacking.

One of the positive aspects of the discussion was that the senior management were aware of the data quality issues and the pressures of compliance, they are highly supportive in improving the current practices and procedures but present organisational culture and remains of public sector heritage is making their task harder and less efficient. The organisation is still developing key metrics or the parameters which can identify the cost of the poor data and poor data decisions

5 CONCEPTUAL FRAMEWORK

Fundamental to this study is the identification of three conceptual elements seen to be key to any data quality programme namely: People, Processes and Data. This has been developed further to form the basis of the conceptual framework (Figure 2) detailed below. This research indicates that there are a myriad of methods and solutions to improve *data* quality in both the areas of transactional and master data at various levels embracing both *process* and *people,* with varying consequences and degrees of success. Nicholaou (2004:44) identified that lack of people training and failure to recognise the effects of an ERP system on current business *processes* are the most important culprits in problematical implementations. Whilst all such initiatives have enormous merit in themselves, they will not generate long-term success or influence unless they can be embedded. This study takes note of these theories and practices that can improve and create quality data, but focuses upon identifying how an organisation may be able to create an environment where data quality improvement initiatives may be sustained. In this it accepts that there must be a climate where such improvements should be sought-after, generated, supported and implemented with adequate resources.

The conceptual framework depicted in Figure 2 above sets Data Quality firmly within the overall context of Data Governance as part of an enterprise-wide data strategy and acting as a route map through the whole research. The initial triple inter-linked framework developed from an intensive review of the literature comprises the 'Data' elements of master data management, together with operational and transactional data; 'Process' review and improvement initiatives running in tandem with the necessary system housekeeping procedures; together with the 'People' elements of education and training, personal development aligned with accessibility in the form of Assistive Technology (hardware and software techniques developed in order to assist visually or physically disabled persons gain access to information technology within the working environment). During the research for this study it became apparent that any enduring improvement is predicated on making lasting changes to both processes and individuals' behaviour and to bring about this, there has to be cultural and organisational change mainly through the interaction of leadership and management at all levels. The framework also identifies how the process of producing quality information derived from quality raw data has parallels with a generic product manufacturing process as discussed above. This useful analogy between a production process and an information system also has strong roots in the literature (Strong, Lee and Wang 1997:104; Wang 1998: 59).

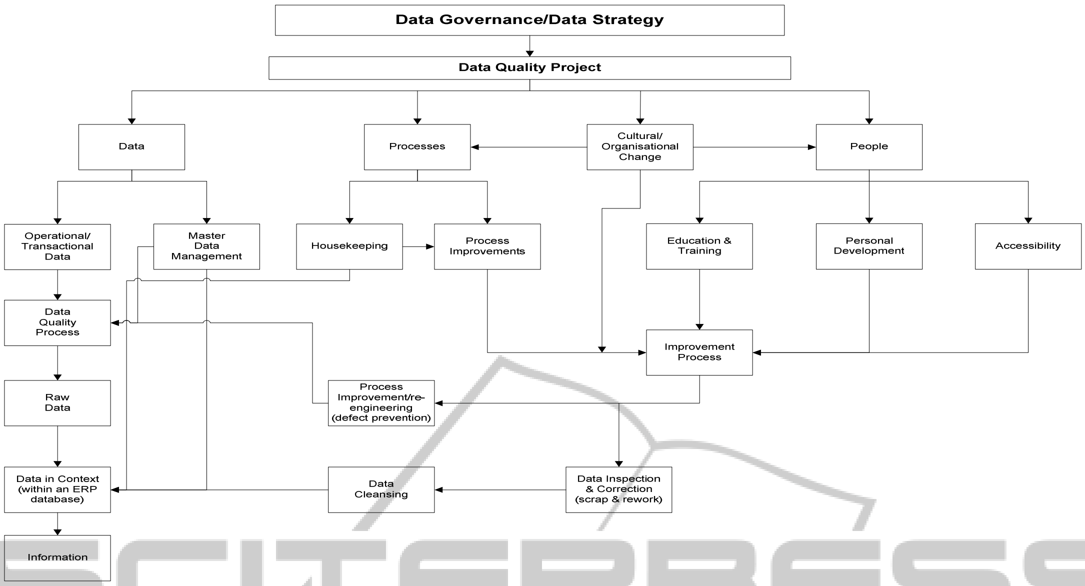

Figure 2: Conceptual Framework

6 CONCLUDING REMARKS

The relationship between theory and practice, research and action to attempt to enhance theoretical knowledge whilst providing practical solutions as expounded by Van de Ven and Johnson (2006) and Van de Ven (2007) has been the overriding ambition of this study. It can therefore be argued that this research has made a contribution to *theory* whilst also assisted in bringing about lasting fundamental practical data quality improvements with real life organisations

REFERENCES

Borek A, Helfert M, Ge M, Parlikad AK. An information oriented framework for relating IS/IT resources and business value. In: *Proceedings of the International Conference on Enterprise Information Systems*. Beijing, China; 2011.

Davenport, T. H. (1998) Putting the Enterprise into the Enterprise System. *Harvard Business Review*, July-August 1998: 121-131.

Davenport, T. H. (2006) Competing on Analytics. *Harvard Business Review*, January 2006: 99-107.

Davenport, T. H. and Harris, J. G. (2002) Elusive data is everywhere: understanding of position and strategy comes before knowledge creation. *Ivey Business Journal*, 66(5): 30-31.

Davenport, T. H. and Harris, J. G. (2007) Competing on Analytics: The New Science of Winning. *Cambridge MA: Harvard Business Schoo*l Press: 22

Davenport, T. H., Harris, J. G., De Long, D. W. and Jacobson, A. L. (2001) Data to Knowledge to Results: Building an Analytic Capability. *California Management Review,* 43(2): 117-138.

Davenport, T. H., 1993. Process innovation: reengineering work through information technology, *Harvard Business Press.*

Dun and Bradstreet and Richard Ivery School of Business. (2006). The Cost of Poor Data Quality: 1-13*: Dun and Bradstreet.*

Eckerson, W. (2002). Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data: 1-33: *The Data Warehouse Institute.*

Economist Intelligence Unit. (2006). Business Intelligence - Putting Information to Work: 25. London.

English, L. P. (2009). Information Quality Applied. Indianapolis: Wiley Publications Inc: 802.

Eppler M, Helfert M (2004) A classification and analysis of data quality costs. 9th MIT *International Conference on Information Quality,* November 5-6, 2004, Boston, U.S.A.

Fehrenbacher, D. and Helfert, M. (2012) "Contextual Factors Influencing Perceived Importance and Trade-offs of Information Quality," *Communications of the Association for Information Systems*: Vol.30, Art. 8.

Fisher, T. (2009). The Data Asset: How Smart Companies Govern Their Data for Business Success. New Jersey: *John Wiley & Sons*: 220.

Folan, P. & Browne, J., 2005. A review of performance measurement: towards performance management. Computers in Industry, 56(7), p.663-680.

Galliers, R. D. and Newell, S. (2001) Back to the Future: From Knowledge Management to Data Management. Paper presented at *the Global Co-Operation in the New Millennium*, Bled, Slovenia.

Ge, M., & Helfert, M. (2007). A Review of Information Quality Research - Develop a Research Agenda. *International Conference on Information Quality*, November 9-11, 2007, Cambridge, Massachusetts, USA

Ge, M. and Helfert M. (2008), Effects of information quality on inventory management. International Journal of Information Quality, 2(2), pp. 176-191.

Gustafsson, P. et al., 2009. Quantifying IT impacts on organizational structure and business value with Extended Influence Diagrams. The Practice of Enterprise Modeling, p.138–152.

Gustavsson, M. and Wanstrom, C. (2009) Assessing Information Quality in Manufacturing Planning and Control Processes. *International Journal of Quality & Reliability Management*, 26(4): 325-340

Harris, J. G. (2005a) Insight-to-Action Loop: Theory to Practice: *Accenture*.

Harris, J. G. (2005b). The Insight-to-Action Loop: Transforming Information into Business Performance: *Accenture*.

Harris, J. G. (2007) Forget the toys- It's the guy with the best data who wins: *Accenture*.

Haug, A., Zachariassen, F., & van Liempd, D. (2011). The cost of poor data quality. *Journal of Industrial Engineering and Management,* 4(2), 168-193.

Hewlett-Packard. (2007). Managing Data as a Corporate Asset: Three Action Steps towards Successful Data Governance: 1-8.

Informatica. (2008). Timely, trusted Data Unlocks the Door to Governance, Risk and Compliance: 15: *Informatica Corporation*.

Information Difference. (2009). The State of Data Quality Today: 33: The Information Difference Company. http://www.pbinsight.com/files/resource-library/resource-files/The-State-of-Data-Quality-Today.pdf

Jung, W. and Olfman, L. (2005), An experimental study of the effects of contextual data quality and task complexity on decision performance, *Information Reuse and Integration Conference,* Las Vegas, Nevada, USA.

Keller, K. L. and Staelin, R. (1987), Effects of quality and quantity of information on decision effectiveness, Journal of Consumer Research, 14(2), pp. 200-213.

Kim W, Choi B (2003) Towards quantifying data quality costs. Journal of Object Technology, 2(4): 69-76

Loshin, D. (2001) The Cost of Poor Data Quality. *DM Review Magazine*, June.

Melville, N., Kraemer, K. & Gurbaxani, V., 2004. Review: Information technology and organizational performance: An integrative model of IT business value. MIS quarterly, p.283–322.

Mooney, J.G., Gurbaxani, V. & Kraemer, K.L., 1996. A process oriented framework for assessing the business value of information technology. *ACM SIGMIS Database*, 27(2), p.68–81.

Mutch, A. (2008) Managing Information and Knowledge in Organizations. London: Routledge: 272.

Orna, E. (2005) Making Knowledge Visible. *Burlington VA: Gower Publishing Company:* 212.

Porter, M. E. & Millar, V., 1985. How information gives you competitive advantage. *Harvard Business Review*, 63(4), p.149–160.

Senge, P. M. (1992) The Fifth Discipline. London: Century Business: 424.

Stenmark, D. (2004) Leveraging Knowledge Management activities in everyday practice. Paper presented at the *CROINFO 2004, Zagreb* Croatia

Strong, D. M., Lee, Y. W. and Wang, R. Y. (1997) Data Quality in Context. *Communications of the ACM,* 40(5): 103-110.

Van de Ven, A. and Johnson, P. E. (2006a) Knowledge for Theory and Practice. *Academy of Management* Review, 31(4): 802-821

Van de Ven, A. H. (2007). Engaged Scholarship: A Guide for Organisational and Social Research. *Oxford: Oxford University* Press: 330

Van Wegen, B. & De Hoog, R. 1996. Measuring the economic value of information systems. *Journal of Information Technology*, 11(3), p. 247-260.

Wang, R. Y. (1998) A Product Perspective on Total Data. *Communications of the ACM,* 41(2): 58-65.