INFORMATION QUALITY: DEFINITIONS, MEASUREMENT, DIMENSIONS, AND RELATIONSHIP WITH DECISION MAKING

Omar A Alshikhi and Bandar M. Abdullah

[1]Quality Analyst, General Authority of Civil Aviation- Saudi Academy of Civil Aviation, Saudi Arabia
[2]Deanship of Scientific Research, University of Jeddah, Saudi Arabia.

ABSTRACT: *Quality data is inevitably an important pre-requisite for managerial decision-making, especially when the decisions made can have far-reaching consequences for the organization. Hence, scrutinizing the information obtained and demanding that the information meets certain features are paramount to achieving sustainable organizational performance. The present paper provides a roadmap on the definition of Quality Data, its dimensions and relationship with decision making effectiveness.*

KEYWORDS: Decision Making, Information, Quality Data.

INTRODUCTION

Business environment is changing all the time. Business is more demanding, and managers must react fast to customers' and competitors' actions. They must make choices and decisions in a complex and uncertain environment. The significance of information as a company's competition factor has been emphasized during the last ten years. Comprehensive and real time information is needed to develop the business.

Decision making processes and their outcomes can be affected by a number of factors. Among them, the quality of the data is critical. Poor quality data cause poor decisions. Although this fact is widely known, data quality (DQ) is still a critical issue in organizations because of the huge data volumes available in their systems.

It has been estimated that up to 5% of data found in organizations are of poor quality (Redman, 2001) and that the average perceived cost of poor data quality is as high as 10% of an organization's revenues (Malcom, 1998). In the healthcare sector, lack of data quality has far-reaching effects. Planning and delivery of services rely heavily on data from clinical, administrative and management sources. For example, evidence-based practice (Strauss et al, 2005) requires access to extensive research data, collated and presented in a way that a clinician can use at the time of diagnosis or in other decision-making situations. The higher the quality of the data, the better will be the patient outcomes. Similarly, quality data, particularly with regard to timeliness and accuracy, are needed for administrative purposes such as hospital bed-rostering and for planning services to ensure that they are cost-effective. These different but interlocking data requirements and decisions ensure that health care organizations and their relationships are inherently complex and demanding (Gendron & D'Onofrio, 2001). Data quality is inextricably linked to the use of Information Systems and the health sector is increasingly an information-driven service (Hovenga et al., 1996). Information held in databases and other electronic repositories and delivered in a reliable and timely manner is critical to the health and well-being of patients, the wider population, and to the management of health care organizations (Long & Seko, 2002). Raising the level of data quality within an organization contributes to improving the quality of decision-making enabling the reduction of uncertainty and the production of more timely and accurate decision outcomes.

Quality Data: Definitions

Ballou and Tayi (1998) define data as 'the raw material for the information age'. A datum is a fact; a value assigned to a variable (Saba & McCormick, 2001), a single observational point that characterizes a relationship (Shortliffe & Barnett, 2000). Data support managerial and professional work and are critical to all decisions at all levels of an enterprise (Fuller & Redman, 1994; Ballou & Tayi 1998). In contrast, information is useful data that have been processed in such a way as to increase the knowledge of the person who uses the data (McFadden et al., 1999). English (1999a) builds on the idea of information being data in context, with knowledge being information in context, where you know the significance of the information. Translating information into knowledge requires personal experience and reflection. Knowledge itself may be processed to generate decisions and new knowledge (Saba & McCormick, 2001) including the results of formal studies and also commonsense facts, assumptions, heuristics (strategic rules of thumb), and models – any of which may reflect the experience or biases of people who interpret the initial data (Shortliffe & Barnett, 2000).Thus, whilst the transformation of data into information is normally an explicit, repeatable and easily conveyed procedure, the further translation of information into knowledge often involves tacit processes that are much more difficult to capture and explain to others (Davidson & Voss, 2002). High-quality data and derived information are needed to create institutional knowledge (stored information) plus reasoning processes that help an organization extract the maximum benefit from the resources. This approach now accepted under the term 'knowledge management' (Davenport, 1998; Davidson & Voss, 2002), draws together the tangible and intangible elements of data and shares them amongst all workers. Data quality is now emerging as a discipline, with specific research programmes underway within universities, the most significant being that of the Sloan School of Management Information Quality Programme at the Massachusetts Institute of Technology (MIT). The field encompasses the well-established Quality Discipline, drawing on the work of Deming (1982), with the adaptation of the plan, do, check, act, cycle' of Crosby (1980), through the notion that 'quality is free' because of the cost of doing things wrong (Juran & Godfrey, 1999).

Further exemplars include the utilization of Six Sigma and Total Quality Management, adapted to Total Data Quality Management (TDQM) and the management of information as a product (Wang et al., 1998).

With the definition of quality explained, there is something to say about data and information quality. On the one hand, both terms are often used interchangeably, and findings are often found to be suitable for both concepts (Helfert et al., 2009). On the other hand, there is a small difference between data and information quality. Literature about data quality often refers to the general definition of quality of Juran (1999). An example is provided by Wang & Strong (1996) who define data quality as "data that are fit for use by data consumers" (p.6). Several authors extended this definition into their research (Tayi & Ballou, 1998; Stvilia et al., 2007). Though, the literature on information quality often follows the quality perspective of meeting customers' expectations (English, 2001; English, 1999; Gustavsson & Wänström, 2009). For instance, Gustavsson and Wänström (2009) define information quality as the "ability to satisfy stated and implied needs of the information consumer" (p. 327). Here customer and consumer of information refer to the user, so the user influences information quality (Naumann & Rolker, 2000).

Print ISSN: 2053-4019(Print), Online ISSN: 2053-4027(Online)

Data and Information Quality dimensions

Although there is a difference between data and information quality, authors agree upon the fact that both are multi-dimensional constructs (Wang & Strong, 1996; Wand & Wang, 1996; Levitin & Redman, 1995). Therefore, the concepts can be split up in broad characteristics, called dimensions, to make them measurable (Sekaran, 2003).

Within literature data and information quality dimensions are described extensively. However, authors refer to the quality dimensions in different ways. Wang and Strong (1996) refer to data quality dimensions, as "a set of data quality attributes that represent a single aspect or construct of data quality" (p. 6). Other authors refer to the same concept, only they refer to information quality dimensions (Gustavsson & Wänström, 2009; Stvilia et al., 2007; Helfert et al., 2009). There are different dimensions described in literature and some are classified, and others are not. A brief grasp in literature provides 13 sets of dimensions, which consist of a different number of dimensions and classifications. An overview of the different dimensions can be found in Appendix A. To select from all dimensions discussed above, the dimensions are counted. Counting of the numbers of authors that refers to a certain dimension, allows for the ranking of the dimensions based on how often they are stated within the different publications. From the 13 publications described, ten stated the dimension of timeliness, nine referred to the dimension of accuracy, eight referred to the dimension of completeness and eight referred to the dimension of relevance. Therefore, the numbers suggest that timeliness, accuracy, completeness and relevance are core dimensions according to the different authors.

Within IS literature, information quality and user satisfaction are two of the major dimensions used for assessing the success of the information systems. According to Wang & Strong, some of the Data Quality (DQ) dimensions that are included within information quality and user satisfaction, are: accuracy, timeliness, precision, reliability, currency, completeness, and relevancy. Some additional dimensions include: accessibility and interpretability. (Wang & Strong, 1996). It is also important to note that the terms DQ and IQ are used interchangeably (Alkhattabi et al., 2010). However, the term IQ is used rather than DQ within contemporary literature due to the following reasons. First off, it has to do with the fact that modern information technology has allowed for information systems to generate not merely data, but also information (Alkhattabi et al., 2010). Since the mobile health apps process the patients' data to generate information to the various stakeholders including healthcare providers, we will for the purpose of this research continue to use the term IQ (Fadahunsi, Forthcoming). In terms of implementing IQ into Health Information Systems, taking into account the importance of IQ in organizations and IS, it can be said the dimensions of IQ is are critical in the context of Health Information Systems, as the health of the community is literally at stake.

According to (Taken from: Laudon and Laudon, 2012): dimensions are the following:

1. Accuracy: is referring to the extent in which data are able to represent reality.
2. Integrity: is defined as the consistency of the structure of data and relationships among the entities and attributes.
3. Consistency: relates to the consistency in the definition of the data elements.
4. Completeness: is about all necessary data being present.
5. Validity: is defined as data values falling within the defined ranges.
6. Timeliness: refers to data being available when needed.
7. Accessibility: is defined as data being accessible, comprehensible and usable.

Print ISSN: 2053-4019(Print), Online ISSN: 2053-4027(Online)

Measuring Data and Information Quality

From the previous section the most important notion is that although data and information quality are different, both are measured using a set of dimensions. In this paragraph the next step to discuss is how these dimensions are measured. Although both data and information have the same dimensions, the way these dimensions are measured is different. The reason for this is that these data and information quality measures refer to respectively objective and subjective assessment of these measures. According to Pipino et al. (2002) objective assessments reflect upon data in databases while subjective assessments reflect the needs and experiences of stakeholders. Also, Helfert et al. (2009) acknowledge this distinction and differentiate between objective and subjective information quality assessment. They state that objective assessment refers to data conforming to quality specifications and references and subjective assessment refers to information being fit to use by information consumers. The distinction between objective and subjective is also supported by Naumann and Rolker (2000) and Forslund (2007). Therefore, in terms of the measurement of the dimensions, there is a difference between data and information quality

Different authors have measured data and information quality, either subjective or objective as being part of their methodology. Often these authors focus on more than just the measurement of data and include for instance also the state reconstruction phase or improvement phase (Batini et al., 2009). An example is the AIMQ methodology by Lee et al. (2002), which is a methodology that has been developed for the assessment and improvement of data quality. However, the authors work with only a subjective assessment and a benchmark on information quality. Others focus on a very specific type of information systems, like Jeusfeld et al. (1998) who present an approach of quality management in a data warehouse. Due to different 'specializations' it is difficult to align a certain methodology to this research. However, the assessment stages in these methodologies do provide a solid foundation for the data and information quality measures used in this research. Therefore, to arrive at a complete set of measures, in total 16 relevant publications on data and information quality are used. From these publications all measures are categorized on their dimension and whether they apply to either data quality (objective measurement) or information quality (subjective measurement).

Quality Information and Decision Making

The most effective format for presenting data-quality information could be a function of the decision-making process or strategy. Over the past several decades a substantial literature has been developed regarding decision making.

The relationship between information quality and decision-making has been labelled complex and therefore has been the subject of extensive research (Slone, 2006). Information quality is considered one of the key determinants for the quality of an organizations' decisions and actions (Stvilia et al., 2007). Therefore, information becomes more often a critical resource to organizations. This is particularly the case, when considering that more than 98 percent of an organizations' assets and those of its customers are managed by data and information (Eckerson, 2002). Lillrank (2003) acknowledges that it is not the primary problem in an organization to do things right, but to have the information that tells what the right things are to do. Experiencing the low quality of information is considered as one of the most serious problems of consumers of information, for both casual users of the web as well as decision makers in an organization (Naumann & Rolker, 2000). Estimates by the Data Warehousing Institute (2002) indicate that poor quality of customer data costs U.S. business $611 billion a

Print ISSN: 2053-4019(Print), Online ISSN: 2053-4027(Online)

year, only on postage, printing and staff overhead. Also, English (1999) tried to assess the business costs of non-quality information and estimated that these costs are around 10 to 25 percent of revenue. In a later article, English (2001) claims that without quality information, an organization cannot thrive. Moreover, problems with information quality can even hurt the organization. Information collected in business processes can add high value, but only if it has quality and is sharable.

Research has shown that including information about the quality of data can impact decision-making (Chengalur-Smith et al., 1999) by enabling decision makers to utilize data more efficiently and effectively (Even et al., 2006). For example, decision makers need to have access to sufficient data quality information to gauge the reliability of the data (Shankaranarayan, et al., 2003). The presentation of the data quality information also can also have an impact (Chengalur-Smith et al., 1999; Ballou & Tayi, 1999). Where simple decision-making is required, Chengalur-Smith et al., found that the participants in their study were prepared to use complex data quality information to assist their choices whereas when faced with complex decisions they resorted to simple metadata. The authors propose that this difference is due to 'information overload' where complex metadata simply compound the complexity of decision-making and make it too difficult. Decisions made in a health care environment are often complex whether due to the multidisciplinary nature of clinical care or the limited resources available for planning and delivering services. However, what is a complex decision to one user may not be complex to another and the use of data quality information is found to increase as experience levels progress from novice to professional (Fisher et al., 2003). Data quality information also assists those who need to make decisions under time pressure, such as in crisis situations (Fisher et al., 2003)

In addition, the past decade presents the upcoming of information systems and technologies that make it possible for managers to use real-time data from the marketplace when making decisions (Laudon & Laudon, 2012). Furthermore, data warehouses are growing and the direct access of information from various sources by information users is improving rapidly (Lee et al., 2002). These three trends increase the need for high-quality information in organizations. Due to the major impact of information quality on organizations, the information quality of the decisions that have the biggest influence on the organization are of special interest. These decisions are the strategic decisions taken in an organization. These strategic decisions are made on the most important drivers of an organization which are derived from the value driver tree (VDT). Because information is critical for an organizations' strategic decision making, this research aims to provide more insight on the relation between information quality and the VDT of EVA. However, in literature this relation has not been addressed before. Even though Brealey et al. (2011) acknowledge that the VDT of EVA is depending on data and therefore is as good as the quality of the data.

CONCLUSION

Organizations are looking for ways to harness the power of big data to improve their decision making. Despite its significance the effects of quality data on decision-making has been given scant attention in the literature. In this paper, the concept of quality data, dimensions, measurement, and links with decision making effectiveness were discussed.

Print ISSN: 2053-4019(Print), Online ISSN: 2053-4027(Online)

REFERENCES

Alkhattabi, M., Neagu, D., & Cullen, A. (2010). Information quality framework for e-learning systems. Knowledge Management & E-Learning: An International Journal (KM&EL), 2(4), 340-362.

Ballou, D. P., & Tayi, G. K. 1999. Enhancing data quality in data warehouse environments, Communications of the ACM, vol. 42 no. 1, pp. 73-78.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methologies for Data Quality Assessment and Improvement. ACM Computing Surveys; 41,3, 16, 1 - 52.

Brealey, R., Myers, S. C., & Allen, F. (2011). Principles of Corporate Finance, Global tenth edition. Irwin, NY: Mc Graw-Hill.

Chengalur-Smith, I. N., Ballou, D. P., & Pazer, H. L. (1999). The impact of data quality information on decision making: an exploratory analysis. IEEE transactions on knowledge and data engineering; 11, 6, 853 – 864.

Crosby, P. 1980 Quality is Free. New York: Penguin Group.

Davenport, T. H. 1998. Working Knowledge: How Organisations Manage What They Know. Boston: Harvard University.

Davidson, C. and Voss, P. 2002. Knowledge Management: An Introduction to Creating Competitive Advantage from Intellectual Capital. Auckland: Tandem Press.

Deming, W. E. 1982. Out of the Crisis. Cambridge: Massachusetts Institute of Technology.

Eckerson, W. W. (2002). Data Quality and the bottom line - achieving business success through a commitment to high quality data. Chatsworth, CA: The Data Warehousing Institute, 101 communications LLC.

English, L. 1999a. Improving Data Warehouse and Business Information Quality. New York: John Wiley & Sons.

English, L. P. (2001). Information Quality Management: The Next Frontier. Quality Congress. ASQ's Annual Quality Congress Proceedings, (pp. 529-533).

Even, A. Shankaranarayanan, G. Watts, S. 2006. Enhancing Decision Making with Process Metadata: Theoretical Framework, Research Tool, and Exploratory Examination. Proceedings of the 39th Hawaii International Conference on System Sciences.

Fisher, C., Chengalur-Smith, I., & Ballou, D. P. 2003. The impact of experience and time on the use of data quality information in decision making, Information Systems Research vol. 14, pp. 170-188.

Forslund, H. (2007). Measuring information quality in the order fulfilment process. International Journal of Quality & Reliability Management; 24, 5, 515-524.

Fuller, B., & Redman, T. C. 1994. Data Quality lessons learned at Telecom Australia, QUALCON 94, pp. 379-386.

Gendron, M. S., & D'Onofrio, M. J. 2001. Data Quality in the Healthcare Industry. Data Quality Journal, vol 7, no.1.

Helfert, M., Foley, O., Ge, M., & Cappiello, C. (2009). Analysing the effect of security on information quality dimensions. ECIS 2009 - 17th European Conference on Information Systems.

Hovenga, E. J. S., Kidd, M. R., & Cesnik, B. 1996. Health Informatics. An Overview . Melbourne: Churchill Livingstone.

Jeusfeld, M., Quiz, C., & Jarke, M. (1998). Design and Analysis of Quality Information for Data Warehouses. Proceedings of the 17th International Conference on Conceptual Modeling (ER'98) (pp. 249 - 362). Singapore: Springer Verlag.

Juran, J. M. (1999). How to Think about Quality. In J. M. Juran, & A. B. Godfrey, Juran's Quality Handbook, Fifth Edition (pp. 2.1-2.18). Irwin, NY: McGraw-Hill.

Juran, J. M., & Godfrey, A. B. 1999. Juran's Quality Handbook (5 ed.). New York: McGraw-Hill.
Laudon, K. C., & Laudon, J. P. (2012). Management Information Systems - Managing the digital firm - 12th edition. London: Pearson Education Limited.
Levitin, A., & Redman, T. (1995). Quality Dimensions of a Conceptual View. Information Processing & Management; 31, 11, 81 - 88.
Lillrank, P. (2003). The quality of information. The International Journal of Quality & Reliability Management; 20, 6/7, 691-703.
Long, J. A., & Seko, C. E. 2002. A new method for database data quality evaluation at the Canadian Institute for Health Information, 7th International Conference on Information Quality. Boston: Massachusetts Institute of Technology.
Malcom, A. (1998). Poor data quality costs 10% of revenues, survey reveals. Viewed 18 Sep 2018 http://computerworld.co.nz/news.nsf/UNID/CC256CED0016AD1ECC25684C000E0278?OpenDocument& Highlight=2,Poor,data,quality,costs
McFadden, F. R., Hoffer, J. A., & Prescott, M. B. 1999. Modern Database Management (Fifth ed.). Reading: Addison-Wesley.
Naumann, F., & Rolker, C. (2000). Assessment Methods for Information Quality Criteria. Proceedings of 5th International Conference on Information Quality, (pp. 148 - 162).
Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Quality Assessment. Communications of the ACM; 45, 4, 211- 218.
Redman, T. C. 2001. Data Quality. The Field Guide. Boston: Digital Press.
Saba, V. K., & McCormick, K. A. 2001. Essentials of Computers for Nurses. Informatics for the New Millennium (Third ed.). New York: McGraw-Hill.
Sekaran, U. (2003). Research Methods for Business - A skill building approach. Fourth Edition. NY, US: John Wiley & Sons, Inc.
Shankaranarayan, G., Ziad, M., & Wang, R. Y. 2003. Managing data quality in dynamic decision environments: an information product approach. Journal of Data Management, vol 14, no. 4, pp. 14-32.
Shortliffe, E. H., & Barnett, G. O. 2000. Medical Data: Their Acquisition, Storage and Use. In L. M. Fagan (Ed.), Computer Applications in Health Care and Biomedicine (2nd ed), pp. 37-69 Springer.
Slone, J. P. (2006). Information quality strategy: an empirical investigation of the relationship beween information quality improvements and organizational outcomes. Minneapolis, United States of America: Capella University.
Strauss, S.E., Scott Richardson, W., Glasziou, P., & Haynes, R.B. 2005. Evidence based medicine (third ed.) London: Churchill Livingstone
Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A Framework for Information Quality Assessment. Journal of the American Society for information science and technology; 58, 12, 1720 - 1733.
Wand, Y., & Wang, R. Y. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM; 39, 11, 86 – 95.
Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems; 12, 4, 5 - 33.
Wang, R. Y., Lee, Y. W., Pipino, L. L., & Strong, D. M. 1998. Manage your information as a product, Sloan Management Review, pp. 95-105.