ON THE PHENOMENON OF INFORMATION DILUTION

Mysore Ramaswamy, Southern University, Baton Rouge, LA, mysore@acm.org

ABSTRACT

Innovations in information technology (IT) have affected all aspects of modern life. The quantity of information available is increasing exponentially. Simultaneously, it is becoming much easier to access these vast amounts of information. Unfortunately, in some cases, this information explosion has also made it harder to get the relevant and accurate information needed for good decision-making. The present day transactional engines generate abundant information, but we still observe problems in decision-making at higher levels of management. There is a need to be wary of sub-optimal decisions based on information culled from the transactional level of management in addition to data quality problems that are inevitable due to the volume of data available. In spite of the tremendous increase in the available information for decision makers, some decisions at the strategic level turn out to be of poor quality. At the macro level, we find senior managers failing to seek answers for critical questions. At the micro level, we have not fully succeeded in monitoring the quality of the components that make the data-information-decision chain. In this paper, we analyze some of the factors that are responsible for this information dilution that is directly responsible for poor quality decision-making. We study the problem from both macro and micro perspectives and propose a domain specific approach to alleviate the problem.

Keywords: Information Dilution, Decision Value, Information Integrity, Data Quality

INTRODUCTION

Ease of access to information and its sheer abundance are some of the characteristics of this information age. According to a report from UC Berkeley's School of Information Management and Systems, during 2002, about five exabytes (1 exabyte = 10^{18} bytes) of new information was produced. Ninety-two percent of the new information was stored on magnetic media, mostly in hard disks. To get a sense of this enormity, we should point out that five exabytes of information is equivalent in size to the information contained in half a million new libraries the size of the Library of Congress print collections. As a direct consequence, assuring the quality of information used in managerial decision-making has become a critical issue. Rockart [10] points out that in a majority of cases, information exists as 'islands,' and these pieces are not linked to help effective decision-making.

In a recent study by the Automotive Aftermarket Industry Association (AAIA) Electronic Commerce Committee, it was estimated that data errors cost $1.7 billion annually for suppliers and distributors. The Internet has provided information access to a degree unimaginable earlier. This abundance of information also makes it imperative that we set some quality standards for information that is used in making decisions as well as data that are used in creating or processing information.

Traditionally, information systems have been designed to cater to the needs of managers at different levels of management. The advent of highly efficient transactional engines such as the ones used in Enterprise Resource Planning (ERP) systems has resulted in the availability of large quantities of information at the transactional level. The managers at the tactical and strategic levels are sometimes apprehensive about using these outputs of hardwired subsystems, as they are not sure of the process logic [3]. For decision-making at higher levels of management, the information required has to be necessarily broad based.

In this paper the term 'information dilution' is used in a general sense with emphasis solely on its effect on the strength of decisions made based on that information and not in the way it is used in the quality assurance literature pertaining to aeronautics. We study the information dilution phenomenon from two different perspectives. First, at a macro level, we look at the concerns that arise out of processing and integration issues and how they are affected by information abundance. Then, at a micro level, the issues pertaining to quality of components of data-information-decision chain are studied. Both these approaches are necessary to analyze the phenomenon of poor quality decision-making in spite availability of abundant information. The rest of the paper is organized as follows: The next section introduces an analysis of issues related to the processing of data/information and integration of data from different sources. Next we address data, information,

and decision quality concerns. The conclusions and future research directions are presented in the last section.

PROCESSING AND INTEGRATION ISSUES

First we look at the issue of lowering decision value in spite of abundant information at a macro level. Bazerman and Chugh [1] refer to the case of Vioxx, and analyze why Merck's senior executives allowed the product to stay on the market for a long time. Merck's CEO insisted that he would have pulled Vioxx earlier had he known the risk it posed. Nobody is complaining of intentional unethical behavior. The lapse was essentially due to poor quality of decision making in spite of availability of abundant information. The fact that 14.6% of Vioxx patients suffered from cardiovascular troubles while taking the drug can be found in Merck's report to federal regulators. This clearly points out to problems in processing and integrating of information that is used in the decision making process.

Rogers and Blenko [11] state the need to strike a balance between global and local considerations in the decision making process. When decisions regarding outsourcing are being made, the problem exacerbates. When pieces of information from different sources are integrated, the manner in which those pieces of information get processed from their respective databases becomes an important issue. This is similar to the problems faced when knowledge bases obtained from different experts are combined together to build large knowledge-based systems. Consistency errors can occur due to (a) data errors, (b) processing anomalies. Data errors can be minimized by strict monitoring of data quality. To address the issue of processing anomalies in different sources, we should adhere to a domain-specific data dictionary.

Another interesting aspect of information dilution is seen when decision makers become reluctant to seek critical information. Bazerman and Chugh [1] also refer to the case of the Challenger disaster and analyze why executives at NASA allowed the launch. Before the launch, there were discussions on whether the combination of low temperature and O-ring failure would be a problem. What was not discussed was the information regarding prevailing temperatures for the previous 17 shuttle launches in which there was no O-ring failure. Information abundance has the potential to give a false sense of confidence to the decision makers and make them less likely to seek critical albeit inconvenient questions.

Contemporary information systems such as enterprise resource planning (ERP) systems attempt to integrate all departments and functions across an organization into a single system that can serve the needs of all those different departments. Theoretically this should lead to cooperation between administrative agencies across organizational boundaries and even across hierarchical echelons [15]. But in many instances, information itself is the primary medium of value and exchange and sharing of such a powerful resource may result in complex mixture of conflict and cooperation between entities [6]. No doubt ERP systems are quite useful at the transactional level. But decision making process at the strategic level doesn't lend itself to integrate with the corporate template of ERP systems.

Integration of information from different organizations can also be problematic due to differing contexts pertaining to policy and social environment, interorganizational setting, business processes, and technology solutions [9]. As information integration can lead to external evaluation and criticism, some organizations may hesitate to cooperate. One way to overcome this barrier is to create practical tools such as metadata inventories and standardized data sharing agreements [4]. Typically, over time information flows tend to change in organizations without corresponding modifications in information systems. This problem accentuates when such systems are integrated. Data definition and semantic translation are critical factors in integration of information from geographically dispersed sources. In order to successfully solve the technical problems, it is necessary to develop domain-specific standards.

Enterprise Resource Planning (ERP) systems have been developed with the twin objectives of reducing operational costs and helping make better quality managerial decision making. 'Work order' is the basic transaction upon which ERP implementations such as SAP have been built. As the production process progresses, the same system will link the consumption of inventory as well as track the production, shipping and accounts receivable [3]. The challenge to the management is to use the available information at the operational level for making decisions at the tactical and strategic levels. Undoubtedly, ERP systems provide vast amounts of information by utilizing their highly efficient transactional engines. The undesirable consequence is the greater degree of complexity in the decision-making process at the higher levels of management due to the increased volume of information available to managers. Because that information has been generated from a 'template' approach and does not

differentiate between different types of data. Clearly, there is a need for an interface between transactional data / information and tactical decision-making.

QUALITY CONCERNS

Now we study the phenomenon of poor decision value from a micro perspective. The value of a decision is directly dependent on the strength of data-information-decision chain. A clear understanding of the quality dimensions along with which information has to be measured is necessary to improve the usefulness of information systems that are used in different levels of management. Current literature on data and information quality considers a lot of dimensions or attributes to measure the quality of input that goes into the decision-making process.

The quality of data plays a dominant role in the final decision-making process. The decision task determines the quality dimensions of data [2]. A data quality problem exists if a given set of reported data cannot provide the information needed for decisions [5]. Dimensionality of data has been well researched [5, 12, 13, 14]. One stream of research has generated data quality attributes by conducting surveys. Even though this methodology is useful in getting a broad perspective on the issue, it is not very helpful in designing modifications to specific information systems to improve their yield quality. Wang [13] refers to data product similar to a tangible product that is produced from raw materials. Managerial decisions exhibit a large degree of diversity when compared to physical products, thereby yielding little help in designing/modifying information systems.

Galway and Hanks [5] categorize data into three groups: operational, conceptual, and organizational. Problems with operational data can occur due to missing data, invalid data, or inaccurate data. Conceptual data errors occur as a result of imprecise or ambiguous definition of data. Errors in organizational data occur due to disconnects between organizations that generate and/or use data.

From the point of view of dimensionality of data, the above comparison broadly points towards accuracy as the main attribute to monitor. Ambiguous definitions or mismatch between data definitions at different organizations finally contribute to data inaccuracy and hence dilution. This is of special concern in the area of outsourcing.

Information is the link between data and decision. Some researchers have used the terms ‘data quality’ and ‘information quality’ interchangeably [2, 12]. Information is the result of processing data. The processing involves data and models. Therefore, in order to precisely identify the quality dimensions, these have to be considered separately. Zmud [16] suggests the following dimensions of information: quantity, reliability, timeliness, and format quality. The dimensionality of information has a direct bearing on how it is used in decision-making process. Clearly, this depends on the level of management at which the decisions are made as well as the specific decision domain for which the information is sought.

Current literature on information quality does not treat the components of data-information-decision chain separately for evaluating quality [2, 7, 8, 12]. We illustrate the need to analyze the components separately by the following example.

Consider the information needed to make a decision regarding production schedule for the subsequent quarter in a manufacturing plant. Quantity to be produced (Q) depends on the following: estimates of demand for various products (ED), inventory of finished goods (IFG), and in-process inventory (IPI). ED in turn, depends on sales data (SLD) of previous time periods as processed by sales managers using appropriate forecasting models in their decision support systems.

Now, let us look at the factors that affect the quality of the parameters mentioned above. Good quality production schedule Q, really refers to a production schedule that uses the resources in an efficient and effective way. What are the dimensions of such a decision? Completeness would imply that all the products to be produced are covered, and relevance refers to efficient and effective use of resources.

ED, the estimates of demands of various products to be manufactured are the outputs of information processing models such as forecasting the inputs for which are SLD, sales data for previous time periods. Quality dimensions of ED are accuracy and objectivity. Here the accuracy of ED also depends upon the appropriateness of the models used in the decision support systems. Quality dimension objectivity is defined to include both the appropriateness of model as well as lack of bias. SLD has to be necessarily accurate. IFG, and IPI have the same quality dimension of accuracy. From the above discussion, we note the domain specific nature of quality dimensions that are appropriate for the different components of data-information-decision chain.

CONCLUSION

The dilution of information in spite of the recent innovations in information technology can be attributed to insufficient data quality, processing anomalies, and problems in integrating information from different sources. The quality of decision gets affected and lowers its value at different levels of management by the quality of data and/or information that are used. In order to improve the quality of any component of data-information-decision chain, it is necessary to clearly conceptualize the dimensionality of those components. This is a necessary precursor for building a quality control mechanism. In this paper, we have studied some of the factors that are responsible for information dilution. At the macro level, there is a need to seek answers to critical questions and not swept over in the deluge of information. At the micro level, we need to monitor the components of the data-information-decision chain so that they can be monitored effectively. We have also suggested a framework where the decision domain is used to define the quality dimensions of the various components of the data-information-decision chain that are used in making that decision.

Future research in this area will consist of building a domain-specific framework that integrates levels of management involved in the decision-making process as well as the different components of data-information-decision chain. Such a framework can be used to yield a monitoring mechanism to safeguard the quality of information given to the users. Future work will also deal with the design of an interface between transactional data and tactical decision-making that takes into account the relative importance of various data clusters at the transactional level and the different criteria for decision-making at the tactical level.

REFERENCES

1. Bazerman, M. H. & Chugh, D. (2006). Decisions without blinders. *Harvard Business Review 84*(1), 88-97, 133.
2. Cai, Y. et al (2003). Evaluating completeness of an information product. *Proceedings of the Ninth Americas Conference on Information Systems,* 2274-2281.
3. Carlton, F. & Adam, F. (2005). Understanding the impact of enterprise systems on management decision making: An agenda for future research. *The Electronic Journal of Information Systems Evaluation, 8*(2), 99-106.
4. Dawes, S. S., Bloniarz , P.A., Kelly, K. L. & Fletcher, P.D. (1999). *Some Assembly required: Building a Digital Government for the 21st Century.* Albany, NY: University at Albany, SUNY.
5. Galway, L. A. & Hanks, C. H. (1996). *Data Quality Problems in Army Logistics.* Santa Monica, CA: Rand.
6. Homburg, V. & Bekkers, V. (2002). The back-office of e-government. *Proceedings of the 35th Hawaii International Conference on Information Systems,* (9 pages).
7. Klein, B. D. (2003). Some determinants of user perceptions of information quality on the World Wide Web. *Proceedings of the Ninth Americas Conference on Information Systems,* 2282-2286.
8. Klein, B. D. (2001). User perceptions of data quality: internet and traditional text sources. *Journal of Computer Information Systems*, *41*(4), 5-15.
9. Pardo, T. A., Cresswell, A. M., Dawes, S. S. & Burke, G. B. (2004). Modeling the Social & Technical Processes of Interorganizational Information Integration. *Proceedings of the 37th Hawaii International Conference on System Sciences,* (8 pages).
10. Rockart, J. (2004). Information: Let's get it right. *MIS Quarterly Executive*, *3*(3), 143-150.
11. Rogers, P. & Blenko, M. (2006). How clear decision roles enhance organizational performance. *Harvard Business Review*, *84*(1), 53-61.
12. Wand, Y. & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86-95.
13. Wang, R. Y., Storey, V. C. & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge & Data Engineering 7*(4), 623-640.
14. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-34.
15. Wimmer, M. R., Traunmuller, R. & Lenk, K. (2001). Electronic business invading public sector: Considerations on change and design. *Proceedings of the 34th Hawaii International Conference on System Sciences, 5,* 1-10.
16. Zmud, R. W. (1978). An empirical investigation of the dimensionality of the concept of information. *Decision Sciences, 9*(2), 187-195.