

December 2004

# A Data Quality Assurance Model in the B2B Networked Environment

Yu Cai

*Boston University*

Ganesan Shankaranarayanan

*Boston University*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2004>

---

## Recommended Citation

Cai, Yu and Shankaranarayanan, Ganesan, "A Data Quality Assurance Model in the B2B Networked Environment" (2004). *AMCIS 2004 Proceedings*. 499.

<http://aisel.aisnet.org/amcis2004/499>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# A Data Quality Assurance Model in the B2B Networked Environment

**Yu Cai**  
Boston University  
ycai@bu.edu

**Ganesan Shankaranarayanan**  
Boston University  
gshankar@bu.edu

## ABSTRACT

Business-to-Business (B2B) electronic commerce environments are characterized by frequent information exchange. The Internet and advanced information technologies (IT) permit organizations to access information<sup>1</sup> across organizational boundaries. It is important to assure organizations of the quality of information they get from other organizations. However, a well-established data quality (DQ) standard does not exist to manage the quality of data exchanged in the B2B networked environment. In this paper, we develop a generic data quality standard for the B2B data exchange by proposing a three-layer model: 1) the “DQ 9000” quality management standard for the information product manufacturing process, 2) standardized data quality specifications in metadata exchanged using XMI, and 3) a third party data quality certification issuer. This model is a first step to systematically address data quality management in the B2B data exchanges.

## Keywords

Business to Business E-commerce, Data Quality, Information Quality, Data Quality Standard, IT infrastructure, Information Product, and Metadata

## INTRODUCTION

Inter-organizational information exchange is an inseparable part of Business-to-Business (B2B) relationship. Information, such as the product description, price, order details, and shipment status, are exchanged in business transactions. In organizational decision-making, relevant information (e.g. the upstream and downstream inventory in the supply chain) gathered from outside organizational boundaries plays an important role. Currently, it appears that most organizations implicitly and naïvely assume the quality of information coming from other organizations is good. How about the reality? Research indicates poor data quality (DQ) is a serious problem within many organizations (Ballou and Tayi, 1999; Kahn, Strong and Wang, 2002; Pipino, Lee and Wang, 2002; Wang, 1998). It is estimated that the cost of data quality problems is 8-25% of an organization’s revenue (or budget in non-profit organizations) (English, 1999; Redman, 1998). Moreover, addressing problems caused by poor data quality could constitute 40-60% of the expenses of a service organization or waste 40-50% of an organization’s IT budget. Based on a data quality survey conducted in 2001, the Data Warehousing Institute (TDWI) estimates that the poor quality customer data alone costs U.S. businesses more than \$600 billion a year (Eckerson, 2002). Although the data quality problem costs billions of dollars each year, the survey found only 11 percent of companies had implemented data quality management programs and 48 percent of all companies still had no plan for managing data quality at that time. Given that the quality of data within most organizations is still far from satisfactory, how can organizations be sure that the quality of data obtained from other organizations is of acceptable quality? The same TDWI data quality survey also illustrates that a major source of data quality problems is from external data.

Large volume and high frequency information exchange as well as a wide variety of data sources characterize the B2B electronic commerce environments. It is important to assure organizations of the quality of information they get from other organizations. Without a common understanding of data quality management across organizations, it is difficult to assess the quality of information exchanged in B2B activities. In this paper, we develop a generic data quality standard for the B2B electronic commerce by introducing a three-layer model. First, to assure the quality of information exchanged in B2B

---

<sup>1</sup> In this paper, following the tradition in “data quality” research, we use the terms “data” and “information” interchangeably

settings, each organization should perform internal data quality management. By taking information product approach (Wang, Lee, Pipipo and Strong, 1998) we propose a “DQ 9000” standard for managing information quality by that providing the guidelines for the manufacturing processes that create information products. Secondly, to facilitate a common and comprehensive understanding of the quality of data coming from other organizations, we develop a detailed data quality specification based on the IPMAP (Shankaranarayanan, Wang and Ziad, 2000). With the ability to exchange IPMAP metadata using XMI (XML Metadata Interchange), an organization can utilize information about the manufacture of an information product it receives from another organization to manage DQ. Thirdly, since the exchanged data in the B2B environment is processed by different organizations, the original owner of the data can manipulate the DQ measures for self-interests. Therefore, the third party data quality assurance agency is proposed to certify the self-reported data quality measures in the B2B information exchange. This model provides a systematic approach to address the data quality of information exchanges in the B2B environment.

The remainder of this paper is organized as follows. Section 2 presents an overview of the relevant literature. Section 3 describes the DQ assurance model for B2B networked environment. The discussion and the future research directions are presented in section 4.

## RELEVANT LITERATURE

### The Value of Inter-organizational Information Exchange

Information has value. Better information can reduce risk and opportunity cost, lead to better decisions and consequently lead to improved efficiency and performance (Banker and Kaufmann, 2004; Hayek, 1945). Inter-organizational information exchange is an essential part of B2B e-business partnerships. First, transaction information needs to be shared (Barrett and Konsynski, 1982). Second, businesses need to get analytical information from other organizations. For instance, analytical information exchanged in a supply chain includes upstream and downstream inventory levels, capacity information, to name a few. Conventionally, firms’ boundaries constrained the accessibility of such information. However, in recent years, sharing analytical information has become a new business practice. One example of analytical information exchange is Procter and Gamble (P&G) and Wal-Mart initiated Vendor-Managed Inventory (VMI) program. P&G can access its products’ point of sale (POS) data and inventory data at different Wal-Mart stores. This helps P&G improve its planning and operational efficiency. In exchange, Wal-Mart gets discounted prices and automatic inventory replenishment from P&G (Ireland and Bruce, 2000; Waller, Johnson and Davis, 1999). Research in operations management has shown that the inter-organizational information sharing helps mitigate the “Bullwhip effect” (Lee, Padmanabhan and Whang, 1997) and contributes to the overall supply chain management performance at the operational level (Cachon and Fisher, 2000; Chen, Drezner, Ryan and Simchi-Levi, 2000; Lee, So and Tang, 2000). The rapid improvement in information technology has dramatically reduced the information processing cost (Malone and Crowston, 1994; Malone, Yates and Benjamin, 1987). IT makes many activities that were impossible, possible. For instance, it is technologically feasible to share the real time POS data along the whole supply chain. With information offering more value as well as lower information processing costs, organizations capture and use more information. This is reflected in B2B information exchanges and research on inter-organizational information exchange has gained significance. This paper addresses one important issue of B2B information exchange in networked environment—assessing the quality of information flowing across organizational boundaries.

In this paper our focus is on the “hard data” which has well defined semantics and structure. Because hard data has objective characteristics that are commonly agreed upon across organizations, it is easier to develop and implement data quality measurements. Some “soft” information, such as new product development plan, is too contextual to have generally agreed quality measurements.

### Data Quality, Information Product Approach and IPMAP

Data quality issues are inherent in information systems management (DeLone and McLean, 1992, 2003). Poor quality data can have severe impacts on organizations’ operation including low customer dissatisfaction, increased cost, and reduced job satisfaction (Redman, 1995, 1998; Wang, 1998).

Several approaches such as data tracking and statistical process control (Redman, 1996), data cleansing (Hernandez and Stolfo, 1998) and Information Product (IP) approach (Wang et al., 1998) have been developed for managing data quality. IP approach views information as product. Wang et al. (1998) defined an IP as the deliverable that corresponds to the specific requirements of end users/consumers. Examples of IPs include invoices, business reports, and prescriptions. The IP approach takes both data and data processing into consideration and provides a foundation for developing systematic frameworks to trace, evaluate and manage data quality in information systems. Using the IP approach we can adopt total quality management (TQM) techniques that have been successfully applied in physical product manufacture to manage data quality. In this paper, we adopt IP approach to develop the data quality assurance model for information exchanged in B2B environments. From the IP perspective, the data/information exchanged across organizational boundaries can be viewed as various types of information products that correspond to users' requirements. The scope of the proposed data quality assurance model is designed generally enough to address those inter-organizational data/information exchanges. A special case of information product exchange in B2B environment is the purchase of information goods. This data quality assurance model is also applicable for facilitating transactions in marketplaces dedicated to information goods.

The information product map (IPMAP) is a graphic model to represent the manufacture of an information product (Shankaranarayanan et al., 2000). It permits the systematic representation of the distribution of the data sources, the flow of data elements and the sequence by which these data elements are processed to create the required IPs within/across organization and information system boundaries. Another important feature of IPMAP is the metadata associated with each construct that includes detailed information such as the flow of data through constructs, the elements of data, and data manipulations performed. IPMAP as a platform independent method can be implemented in any organization for DQ management. Moreover, the IPMAP metadata can be shared with other organizations to provide a comprehensive understanding of the quality of data exchanged across B2B boundaries. Because of those features, IPMAP provides a powerful vehicle to evaluate data quality, and hence we adopt the IPMAP for the proposed data quality standard.

#### DATA QUALITY ASSURANCE MODEL

In B2B networked environment, players interact with numerous different business partners. Information is processed in each organization's own proprietary information systems before it is exchanged across organizational boundaries. Without a common understanding of data quality management across organizations, it is difficult to assess the quality of information exchanged in B2B activities. To develop a data quality standard for the B2B networked environment, three requirements must be met.

1. Organizations must perform internal data quality management. (*They know how to speak.*)
2. Organizations adopt the use of IPMAP as instrument and a similar set of metrics to evaluate each DQ dimension. (*They speak the same language.*)
3. In the B2B data exchange relationship, players agree to share the relevant IPMAP metadata (which includes the DQ data) associated with the exchanged data. (*They talk to each other and understand each other.*)

Based on the nature of the information exchanged in the B2B networked environment, we propose the following three-layer data quality assurance model by adopting the IP approach.

1. General data quality management standards "DQ 9000".
2. Detailed data quality specifications in IPMAP metadata.
3. Third party data quality assurance certification.

By viewing information as products, the proposed framework adopts the well-accepted concept from the physical product quality management practice. In the following example by examining a physical product's quality, we illustrate the basic concepts in the three-layer data quality assurance model.

*To protect against the infection of Severe Acute Respiratory Syndrome (SARS) virus, a high quality mask is required. But how do we ensure that a specific mask is qualified? As consumers, we can interpret the product quality from different perspectives. For instance, in a mask's product description, we find the following information: the mask is manufactured by a company complying with the ISO 9000 standards; it is certified by the National Institute for Occupational Safety and Health (NIOSH) as a type N95 respirator and recommended by the World Health Organization (WHO) as a SARS prevention mask. After further research, we find that the "N95" specification means the mask is capable of filtering at least 95% of the 0.3µm particles (non-oil containing dusts, fumes and mists), and it can last for 4 hours in regular use.*

The ISO 9000 complying information tells us the manufacturer has implemented a set of well-established quality management standards to control product quality, which implies the mask comes out of a reasonably good quality production process. This is the analogous to the general data quality standards "DQ 9000". The NIOSH approval and WHO recommendation information indicate this product is certified by third party organizations as meeting some quality standards. This is analogous to the third party data quality assurance certification. The last piece of information about the technical specification tells us the detailed quality standard it meets. This is analogous to the detailed data quality specification in IPMAP metadata. As long as we have the above three pieces of information about a product (and assuming the information are authentic), we can understand the quality of that product. Following the same logic, we propose a three-layer data quality assurance model for information exchange in the B2B networked environment.

### **"DQ 9000" Quality Standard**

The ISO 9000 family is a set of well established, high-level, generic and auditable standards which specify quality management system requirements for use when a contract between two parties requires the demonstration of a supplier's capability to design and supply product (Paulk, 1995). It is sponsored by the International Organization for Standardization and enjoys global prestige. "ISO 9000 has become an international reference for quality management requirements in business-to-business dealings..." (www.iso.ch). ISO 9000 provides a conceptual framework for improving the quality of products and services in a disciplined way. Organizations typically use ISO 9000 standards to regulate their internal quality system (ensuring control, repeatability and good documentation of processes) and assure the quality system of their suppliers.

Following the success of ISO 9000 for physical products and services, we propose developing and implementing a similar set of standards, "DQ 9000", for information product quality management. DQ 9000 is a generic standard for DQ management in all business processes. The key elements of DQ 9000 include 1) process description; 2) record keeping; 3) administrative responsibility; and 4) dedicated DQ assurance unit. DQ 9000 provides guidelines to precisely document the whole information product manufacturing process. It requires the detailed description of each information processing stage including data collection, data manipulation, and data storage. Organizations need to keep accurate records on the source of data, the location of data, data attributes, the quality of data along multiple data quality dimensions such as accuracy, timeliness, and completeness, data processing specifications, data use, person/role responsible for data processing/storage, and the time when these activities occur. DQ 9000 outlines the administration responsibility of data quality management. It calls for organizations to develop and implement high-level data quality policies, explicit data quality documentation, and comprehensive data quality operating procedures/working instructions (e.g. what should do with the missing values; how regularly the database/data warehouse should be updated). After clearly defining how information is processed in all settings, organizations must make sure they strictly follow those rules in practice. DQ 9000 also requires organization to establish dedicated DQ assurance unit to oversee DQ management.

The DQ 9000 borrows the essence of process quality management from ISO 9000. Previous DQ management methods (Hernandez and Stolfo, 1998; Redman, 1996) are mostly "data driven" and based on techno-centric logic. Just like IP approach, the DQ 9000 has the unique advantage that it considers both the "tree" (data) and the "forest" (business process/business eco-system) simultaneously. It may be argued that implementing DQ 9000 is only about "doing things right" in information product manufacturing, which does not guarantee data quality. But doing things right in processes is a necessary condition to achieve the desired outcome. For example, a national retail company's service division implemented a mobile communication system that integrated the data quality assurance procedures into its business processes (e.g. service representatives are enforced to input the part/assembly used into the system before issuing the invoice to the customer. The system automatically updates part/assembly inventory after each service trip, so that part/assembly can be correctly

replenished before next service visit). The adequate data entering process improves the data quality. And it greatly increases the inventory management efficiency and the customer satisfaction level. It is estimated to save the company more than 1 billion dollars a year (Henderson, 2003). This illustrates the benefit of incorporating data quality management within processes.

DQ 9000 is intentionally designed to be general enough to fit all kinds of organizations. Under DQ 9000 guidelines, organizations can choose various DQ management tools/techniques to meet their specific requirements. We believe developing and implementing a generic DQ 9000 quality management standards will systematically improve data quality management process within organizations. When most organizations implement “DQ 9000”, the collective quality of all IPs will increase. As a result, the quality of overall information exchanges in the B2B networked environment will be improved. This defines our first requirement: Organizations perform internal data quality management.

### The Data Quality Specification in IPMAP Metadata

IPMAP is generic and can be used in any organization for data quality management. To facilitate the capture and storage of IPMAP metadata, we propose the following definitions in Backus Naur Form (BNF) for the IPMAP metadata:

```

IPMAP Metadata ::= [Block info] [Data info]
Block info ::= [Block name] [Block type] [Block physical location] [Business unit responsible] [Block description]
              [Data quality dimensions]
Data info ::= [Data ID] [Data source] [ Data destination ] [Data type] [Data quality measurements] [Business
              logic/rules]
Block name ::= {String}
Block type ::= source block | the processing block | the data quality check block | the organizational boundary block
              | the information system boundary block | the sink block
Block physical location ::= {String}
Business unit responsible ::= {String}
Block description ::= {String}
DQ dimension ::= {Accuracy, Timeliness, Completeness, other DQ dimensions}
Data ID ::= {String}
Data source ::= {Block name}
Data destination ::= {Block name}
Data type ::= {String}
DQ measurement ::= {floating point number}

```

The IPMAP metadata includes the structure of the IPMAP (what are the manufacturing stages and how information flows through these stages in the IPMAP). Given this metadata, we can construct the IPMAP. IPMAP metadata also provides detailed data quality specifications on how data quality is measured along well-defined quality dimensions. Currently, we include three such data quality dimensions--accuracy, timeliness and completeness in the DQ measurement. Though several dimensions have been defined in the literature for managing data quality we include these three for the following reasons: first these three dimensions are well-addressed in DQ literature (Fox, Levitin and Redman, 1994, Miller, 1996; Wand and Wang, 1996; Wang and Strong, 1996); and second, accepted methods for evaluating data quality along these dimensions have been defined (Ballou and Pazer, 1995; Ballou, Wang, Pazer and Tayi, 1998; Cai, Shankaranarayanan and Ziad, 2003; Shankaranarayanan, Ziad and Wang, 2003). (For the detailed measurement framework, please refer to Appendix 1.) Extensible Markup Language (XML) is a simple and platform independent data description language. The IPMAP metadata can be defined using Document Type Definition (DTD) or schema in XML, which includes the format of the IPMAP metadata tags, elements and attributes. Businesses can use IPMAP and DQ measurement to manage the quality of internal data.

In B2B information exchange, if the IPMAP metadata is transferred along with the information product across organizational boundary, the receiving party can have a better understanding of the quality of the data. Therefore, the third requirement states that in the B2B data exchange relationship, players agree to share the relevant IPMAP metadata (which includes the DQ data) associated with the exchanged data. To facilitate distributed systems to share metadata over the network, the XML Metadata Interchange (XMI) is developed. XMI provides a standard way to exchange information about metadata between modeling tools and metadata repositories. When organizations exchange data in B2B networked environment, they would also transfer the IPMAP metadata associated with the information product via XMI. The receiving organization can reverse-engineer the metadata to recreate the IPMAP that illustrates how the sending organization processed data to create the IP. It can further integrate the IPMAP with one or more IPMAPs of their own. Based on the integrated IPMAP, the accuracy, timeliness and/or completeness can be calculated at any stage of IPMAP to evaluate the quality of the IP. Further more, if a data quality problem is identified at some stage in the IPMAP, the IPMAP metadata offers the ability to trace back and pinpoint the stages that may have caused this and offers the ability to look ahead and target the stages and IP(s) that are likely to be affected by this problem (Shankaranarayanan et al., 2003).

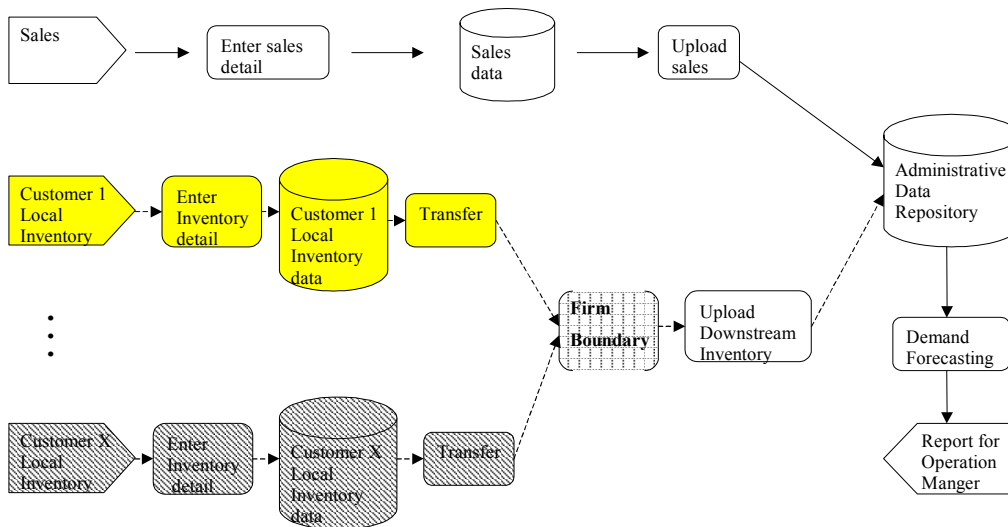


Figure 1. The integrated IPMAP across firm boundary

Figure 1 illustrates how the IPMAP metadata standard works. It is a simplified IPMAP for a firms' demand forecasting process. In supply chain management, to do demand forecasting, ideally a firm needs both past sales data and the downstream customers' inventory level data. But the downstream inventory data is beyond the firm boundary. If all organizations use the IPMAP to manage DQ and downstream customers (customer 1 to customer X) agree to share their inventory level information along with its corresponding IPMAP (metadata), a firm can integrate its downstream customers' IPMAPs into its local IPMAP (linked at the "firm boundary" block). This integrated IPMAP provides a much more comprehensive view of data sources and associated data quality. Now the firm can use both local sales data and downstream inventory data for demand forecasting. Furthermore, it has detailed DQ specifications of downstream customers' inventory data and can use them to evaluate DQ at all different stages.

### The Third Party DQ Certification

The purpose of the third party quality assurance is authenticating data quality in the B2B networked environment. Since the exchanged data in the B2B networked environment are processed by different organizations, the original owner of the data has the potential to manipulate the DQ measures. An effective method to assure the data quality specifications received from other organizations is the third party assurance mechanism.

We further propose a third party DQ auditor for the B2B DQ assurance model. The role of the DQ auditor is like the role of the “Verisign” ([www.verisign.com](http://www.verisign.com)) in data security -- an independent and neutral entity. We expect that it will operate in the following manner: first, the DQ auditor checks whether the organization to be certified uses the general accepted DQ management standard and the IPMAP representation. Then the DQ auditor collects the exchanged IPMAP metadata in XML to test for the potential violation of the IPMAP definition and rules. Second, the DQ auditor conducts the initial DQ auditing for a set of data in that organization to check the validity of the organization’s self-reported DQ measures. If the organization passes the initial test, depending upon the quality dimension and the conformance, the auditor may certify the IP at corresponding levels. For instance, if the organization can only update the inventory data weekly, the quality of inventory data on timeliness dimension is certified at the weekly level. If the organization can update the inventory at the daily level, the quality of inventory data on timeliness dimension is certified at higher daily level. Finally, the DQ auditor conducts random audits to ensure the consistency. Figure 2 illustrates the overall DQ assurance model in B2B networked environment.

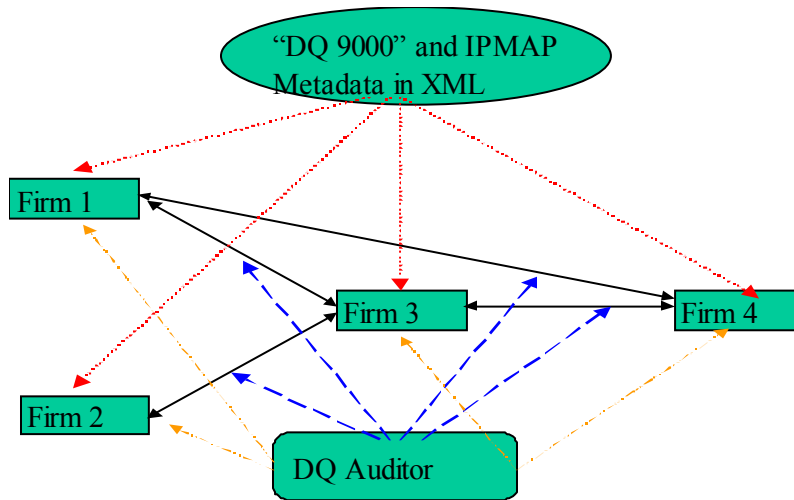


Figure 2. The data quality assurance model in B2B networked environment

Upon request, the DQ auditor will issue DQ certificate. Of course, just like other third party auditing services, DQ auditor cannot prevent cheating. The organization holds the final responsibility for DQ management.

## DISCUSSIONS AND FUTURE RESEARCH

B2B electronic commerce environments are characterized by frequent information exchange. The Internet and advanced information technologies permit organizations to access information across organizational boundaries. It is important to assure organizations of the quality of information they get from other organizations. However, a well-established data quality standard does not exist to manage the quality of data exchanged in the B2B networked environment. In this paper, we develop a generic data quality standard for the B2B data exchange by proposing a three-layer model: the “DQ 9000” quality management standard for the information product manufacturing process, a standardized data quality specification metadata exchanged using XMI, and a third party data quality certification mechanism. This framework is a first step to systematically address data quality management in the B2B networked environment. Currently this standard is still at the theoretical stage. We are examining the issues that need to be addressed for implementing this standard.

One important issue is the burden imposed on the third party “DQ” certificate issuers. Considering the huge volume of information exchanged in the B2B networked environment, it is a tremendous amount of work for the third party DQ certificate issuers to check the quality of all IPs. We expect the evolution of the third party certificate issuers to be a gradual process. At the beginning, some big organizations identify the need to assure the quality of information coming from other organizations. They require companies doing business with them to provide high quality information meeting specified



criteria. For cost efficiency, those big companies outsource the DQ verification task to other organizations (the third party DQ auditors). In order to pass the DQ requirements and gain trust from big companies, some small companies have to get the third party DQ auditor's certification. Furthermore, the third party certificated issuers' responsibility should be like CPA firms. Its job is doing due diligence work to ensure that the client's statement about its data quality is accurate. The ultimate responsibility to meet the data quality requirement rests with the client.

Currently we are working on two directions following this research. One is developing new DQ dimension measurements and refining existing DQ dimension measurement metrics to enhance the proposed framework. The other is trying to integrate the contextual concerns into the measurement construct to make the framework flexible enough to fit in idiosyncratic and dynamic settings.

Research has repeatedly illustrated that the technology itself is not the "silver bullet". The adoption and success of new technology/system/solution are significantly influenced by the dynamic interactions among different stakeholders. In the proposed B2B data quality assurance model, the data supplier, the data customer/user, the third party DQ certificate issuer, and the IPMAP integration software provider have different interests and expectations. Further research from the stakeholder perspective is necessary to ensure successful adoption of this model.

## REFERENCES

1. Ballou, D. P. and Pazer, H. L. (1995) Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff, *Information Systems Research*, 6, 1, 51-72.
2. Ballou, D. P. and Tayi, G. K. (1999) Enhancing data quality in data warehouse environments, *Communications of the ACM*, 42, 1, 73.
3. Ballou, D. P., Wang, R. Y., Pazer, H. and Tayi, G. K. (1998) Modeling information manufacturing systems to determine information product quality, *Management Science*, 44, 4, 462-484.
4. Banker, R. D. and Kaufmann, R. J. (2004) The Evolution of Research on Information Systems: A Fiftieth-Year Survey of the literature in Management Science, *Management Science*, 50, 3, 281-289.
5. Barrett, S. and Konsynski, B. (1982) Inter-Organization Information Sharing Systems, *MIS Quarterly*, Special issue, 93-105.
6. Cai, Y., Shankaranarayanan, G. and Ziad, M. (2003) In *The Proceedings of Ninth Americas Conference on Information Systems (AMCIS)*, Tampa, 2274-2281.
7. Cachon, G. P. and Fisher, M. (2000) Supply Chain Inventory management and the value of shared Information, *Management Science*, 46, 8, 1032-1048.
8. Chen, F., Drezner, Z., Ryan, J. K., and Simchi-Levi, D. (2000) Quantifying the bullwhip effect in a simple supply chain: the impact of forecasting lead times, and information, *Management Science*, 46, 3, 436-443.
9. DeLone, W. H. and McLean, E. R. (1992) Information Systems Success: The Quest for the Dependent Variable, *Information Systems Research*, 3, 1, 60-95.
10. DeLone, W. H. and McLean, E. R. (2003) The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, *Journal of Management Information Systems*, 19, 4, 9-30.
11. Eckerson, W. W. (2002) *Data Quality and the Bottom Line*, The Data Warehousing Institute, Seattle, WA.
12. English, L. (1999) *Improving Data Warehouse and Business Information Quality*, John Wiley & Sons, NY.
13. Fox, C., Levitin, A. and Redman, T. C. (1994) The notion of data and its quality dimensions, *Information Processing & Management*, 30, 1, 9-19.
14. Hayek, F. (1945) The Use of Knowledge in Society, *American Economic Review*, 35, 4, 3-17.
15. Henderson, J. (2003) Boston University BUILDE working paper.

16. Hernandez, M. A. and Stolfo, S. J. (1998) Real World Data is Dirty: Data Cleansing and the Merge/Purge Problem, *Journal of Data Mining and Knowledge Discovery*, 2, 1, 9-37.
17. Ireland, R. and Bruce, R. (2000) CPFR Only the Beginning of COLLABORATION, *Supply Chain Management Review*, 4, 4, 80-87.
18. Kahn, B. K., Strong, D. M. and Wang, R. Y. (2002) Information quality Benchmarks: Product and Service Performance, *Communications of the ACM*, 45, 4, 184-193.
19. Lee, H.L., Padmanabhan, V. and Whang S. (1997) Information Distortion in a Supply Chain: The Bullwhip Effect, *Management Science*, 43, 4, 546-558.
20. Lee, H. L., So, K. C. and Tang, C. S. (2000) The value of information sharing in a two-level supply chain, *Management Science*, 46, 5, 626-643.
21. Malone, T. W. and Crowston, K. (1994) The Interdisciplinary Study of Coordination, *ACM Computing Surveys*, 26, 1, 87-119.
22. Malone, T. W., Yates, J. and Benjamin, R. I. (1987) Electronic Markets and Electronic Hierarchies, *Communications of the ACM*, 30, 6, 484-497.
23. Miller, H. (1996) The multiple dimensions of information quality, *Information Systems Management*, 13, 2, 79-82.
24. Paulk, Mark C. (1995) How ISO 9001 Compares With the CMM, *IEEE Software*, January 1995, 74-83.
25. Pipino, L. L., Lee, Y. W. and Wang, R. Y. (2002) Data Quality Assessment, *Communications of the ACM*, 45, 4, 212-218.
26. Redman, T. C. (1995) Improve Data Quality for Competitive Advantage, *Sloan Management Review*, 37, 2, 99-107.
27. Redman, T. C. (Ed.) (1996) *Data Quality for the Information Age*, Artech House, Boston, MA.
28. Redman, T. C. (1998) The impact of poor data quality on the typical enterprise, *Communications of the ACM*, 41, 2, 79-82.
29. Shankaranarayanan, G., Wang, R. and Ziad, M. (2000) IPMAP: Representing the Manufacture of an Information Product In *The Proceedings of MIT Data quality Conference (IQ 2000)* Boston, MA.
30. Shankaranarayanan, G., Ziad, M. and Wang, R. Y. (2003) Managing Date Quality in Dynamic Decision Environment: An Information Product Approach, *Journal of Database Management*, Vol. 14, 14-32.
31. Waller, M., Johnson, M. and Davis, T. (1999) Vendor-managed Inventory in the retail supply Chain, *Journal of Business Logistic*, 20, 1, 183-203.
32. Wand, Y. and Wang, R. Y. (1996) Anchoring data quality dimensions in ontological foundations, *Communications of the ACM*, 39, 11, 86-95.
33. Wang, R. Y. (1998) A product perspective on Total Data Quality Management, *Communications of the ACM*, 41, 2, 58-65.
34. Wang, R. Y., Lee, Y. W., Pipino, L. L. and Strong, D. M. (1998) Manage Your Information as a Product, *Sloan Management Review*, 39, 4, 95-105.
35. Wang, R. Y. and Strong, D. M. (1996) Beyond accuracy: What data quality means to data consumers, *Journal of Management Information Systems*, 12, 4, 5-34.

## APPENDIX 1 THE TIMELINESS, ACCURACY AND COMPLETENESS MEASUREMENT

### Timeliness

- The extent to which information is sufficiently up-to-date for the task at hand
  - $Timeliness = \{MAX [(1-Currency/Volatility), 0]\}^S$  where
    - $Currency = (Delivery\ Time - Input\ Time) + Age$
    - $Volatility = \text{“Shelf Time”}$
    - $S$ : Parameter to control the sensitivity
- Ballou et al. (1998)

### Accuracy

- The degree to which the reported value is in conformance with the actual or true value. (Ballou and Pazer, 95)
  - $Accuracy = 1 - (\# \text{ of data items in error} / \text{Total } \# \text{ of data items})$
  - $Accuracy = 1 - \{(Correct\ Value - Actual\ Value\ Used) / Correct\ Value\}$
- Shankar et al. 2003

$$A_S = [\sum_{i=1,n} (a_i * A_i)] / [\sum_{i=1,n} (a_i)]$$

- $A_i$  denote the specified accuracy of raw data element  $i$
- $a_i$  decision-maker's perceived accuracy of the data element  $i$

### Completeness (Cai et al. 2003)

- Completeness measurement using a purely objective approach.
  - $D_j$ : The completeness of the  $j$ th raw data element.  
Let  $D_j = 0$ , if the value is missing  
 $D_j = 1$ , if the value is present
  - $E_i$ : The completeness of the  $i$ th IP component.

$$E_i = \frac{\sum_{j=1}^m D_j}{m}$$

$m$  = the number of raw data points needed to get  $E_i$

- Let  $C$  define the completeness measure of an IP.

$$C = \sum_{i=1}^n w_i E_i$$

$N$  = The number of IP components that make up the final IP

$W_i$  : The weight of  $E_i$  in Information Product  $C$ ; ( $\sum W_i = 1$ )