CHAPTER 1

Introduction

Many important corporate initiatives, such as Business-to-Business commerce, integrated Supply Chain Management, and Enterprise Resources Planning are at a risk of failure unless data quality is seriously considered and improved. In one study of a major manufacturer, it was found that 70% of all orders had errors. Aspects of this problem have even gained the attention of the *Wall Street Journal,* which reported that "A major financial institution was embarrassed because of a wrong data entry of an execution order of $500 million..." "Some Northeast states are filing multi-million dollar suits against TRW for bad credit data..." "Mortgage companies miscalculated home owners' adjustable rate monthly mortgage payments, totaling multi-billion dollars..." The major news networks have also reported that many patients died or became seriously ill because of prescription data errors, prompting the Clinton Administration to issue an executive order to address the problem. Anecdotes of high-stake data quality problems abound in the academic literature and news articles.

The field of data quality has witnessed significant advances over the last decade. Today, researchers and practitioners have moved beyond establishing data quality as a field to resolving data quality problems, which range from data quality definition, measurement, analysis, and improvement to tools, methods, and processes. As a result, numerous data quality resources are now available for the reader to utilize [10, 29]. Many professional books [11, 12, 19, 24, 26], journal articles, and conference proceedings have also been produced. Journal articles have been published in fields ranging from Accounting [6, 17, 21, 36] to Management Science [1, 5] to Management Information Systems [2, 25, 35] to Computer Science [3, 4, 13, 16, 18, 20, 23, 27, 31, 34], and Statistics [7, 14, 22]. For example, Communications of the ACM, a premier publication of the Association of Computing Machinery, featured a special section entitled "Examining Data Quality" [28].

Not-for-profit academic conferences and commercial conferences on data quality are part of today's data quality landscape. Back in 1996, the Total Data Quality Management (TDQM) program at the Massachusetts Institute of Technology pioneered the first academic conference for exchanging research ideas and results between researchers and practitioners. The overwhelmingly positive feedback from the conference participants led the organizers to hold the conference every year since. Today, it is a well-established tradition to hold the *MIT Conference on Information Quality* annually. Additionally, many information systems and computer science conferences now feature tracks or tutorials on data quality. Commercial conferences include those organized by Technology Transfer Institute [30] and others. These commercial conferences typically feature invited speakers as well as vendor and consultant workshops. In short, the field of data quality has evolved to such a degree that an exposé of research and practice of the field is both timely and valuable.

FUNDAMENTAL CONCEPTS

Before we present the theme of this book, we first introduce some fundamental concepts that we have found to be useful to researchers and practitioners alike.

Data vs. Information

Data and information are often used synonymously in the literature. In practice, managers intuitively differentiate between information and data, and describe information as *data* that have been processed. Unless specified otherwise, this book will use the term data interchangeably with the term *information.*

Product vs. Information manufacturing

An analogy exists between quality issues in product manufacturing and those in information manufacturing, as shown in Table 1.1. Product manufacturing can be viewed as a processing system that acts on raw materials to produce physical products. Analogously, information manufacturing can be viewed as a system acting on raw data to produce information products. The field ofproduct manufacturing has an extensive body of literature on *Total Quality Management* (TQM) with principles, guidelines, and techniques for product quality. Based on TQM, knowledge has been created for data quality practice.

Table 1.1 : Product vs. Information manufacturing

	Product Manufacturing	Information Manufacturing
Input	Raw Materials	Raw Data
Process	Assembly Line	Information System
Output	Physical Products	Information products

(Source: IEEE Transactions on Knowledge and Data Engineering [34])

An organization would follow certain guidelines to scope a data quality project, identify critical issues, and develop procedures and metrics for continuous analysis and improvement. Although pragmatic, these approaches have limitations that arise from the nature of raw materials, namely data, used in information manufacturing. Whereas data can be utilized by multiple consumers and not depleted, a physical raw material can only be used for a single physical product. Another dissimilarity arises with regard to the intrinsic property of *timeliness* of data. One could say that a raw material (e.g., copper) arrived just in time (in a timely fashion); however, one would not ascribe an intrinsic property of timeliness to the raw material. Other dimensions such as the *believability* of data simply do not have a counterpart in product manufacturing [5].

The Information Manufacturing System

We refer to an information manufacturing system as a system that produces information products. The concept of an information product emphasizes the fact that the information output from an information manufacturing system has value transferable to the consumer, either internal or external. We define three roles (called the three C's) in an information manufacturing system:

1. Information ***collectors*** are those who create, collect, or supply data for the information product.
2. Information ***custodians*** are those who design, develop, or maintain the data and systems infrastructure for the information product.
3. Information ***consumers*** are those who use the information product in their work.

In addition, we define information product managers as those who are responsible for managing the entire information product production process and the information product life cycle. Each of the three C's is associated with a process or task

- Collectors are associated with data-production processes
- Custodians with data storage, maintenance, and security
- Consumers with data-utilization processes, which may involve additional data aggregation and integration.

We illustrate these roles with a financial company's client account database. A broker who creates accounts and executes transactions has to collect, from his clients, the information necessary for opening accounts and executing these transactions, and thus is a collector. An information systems professional who designs, develops, produces, or maintains the system is a custodian. A financial controller or a client representative who uses the information system is a consumer. Finally, a manager who is responsible for the collection, manufacturing, and delivery of customer account data is an information product manager.

Information Quality is a Multi-Dimensional Concept

Just as a physical product (e.g., a car) has quality dimensions associated with it, an information product also has information quality dimensions. Information quality has been viewed as *fitness for use by information consumers,* with four information quality categories and fifteen dimensions identified [35]. As shown in Table 1.3, intrinsic information quality captures the fact that information has quality in its own right. *Accuracy* is merely one of the four dimensions underlying this category. Contextual information quality highlights the requirement that information quality must be considered within the context of the task at hand. Representational and accessibility information quality emphasize the importance of the role of information systems.

Table 1.3: Information Quality Categories and Dimensions

Information Quality Category	Information Quality Dimensions
Intrinsic Information Quality	Accuracy, Objectivity, Believability, Reputation
Accessibility Information Qulity	Accessibility, Access eecurity
Contextual Information Quality	Relevancy, Value-Added, Timeliness, Completeness, Amount of information
Representational Information Quality	Interpretability, Ease of understanding, Concise representation, Consistent representation, Ease of manipulation

(Source: Journal of Management Information Systems [35])

The TDQM Cycle

Defining, measuring, analyzing, and continuously improving information quality is essential to ensuring high-quality information products. In the TQM literature, the widely practiced Deming cycle for quality enhancement consists of: Plan, Do, Check, and Act. By adapting the Deming cycle to information manufacturing, we have developed the TDQM cycle [32], which is illustrated in Figure 1.1.

The definition component of the TDQM cycle identifies important data quality dimensions [35] and the corresponding data quality requirements. The measurement component produces data quality metrics. The analysis component identifies root causes for data quality problems and calculates the impacts of poor quality information. Finally, the improvement component provides techniques for improving data quality. These components are applied along data quality dimensions according to requirements specified by the consumer.

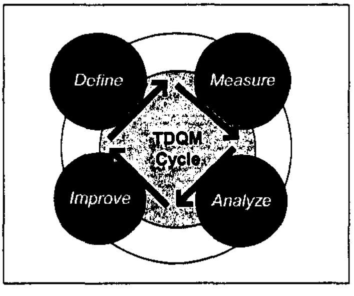

Figure 1.1: The TDQM Cycle

(Source: Communications of the ACM [32])

A FRAMEWORK FOR TDQM

In applying the TDQM framework, an organization must: (1) clearly articulate the information product (IP) in business terms; (2) establish an ***IP team*** consisting of a senior manager as a champion, an IP engineer who is familiar with the TDQM framework, and members from information collectors, custodians, consumers, and IP managers; (3) teach data quality assessment and data quality management skills to all the IP constituents; and (4) institutionalize continuous IP improvement.

A schematic of the TDQM framework is shown in Figure 1.2. The tasks embedded in this framework are performed in an iterative manner. For example, an IP developed in the past may not fit today's business needs for private client services. This should have been identified in the continuous TDQM cycle. If not, it would be the IP team's responsibility to ensure that this need is met at a later phase; otherwise this IP will not be fit for use by the private client representatives,

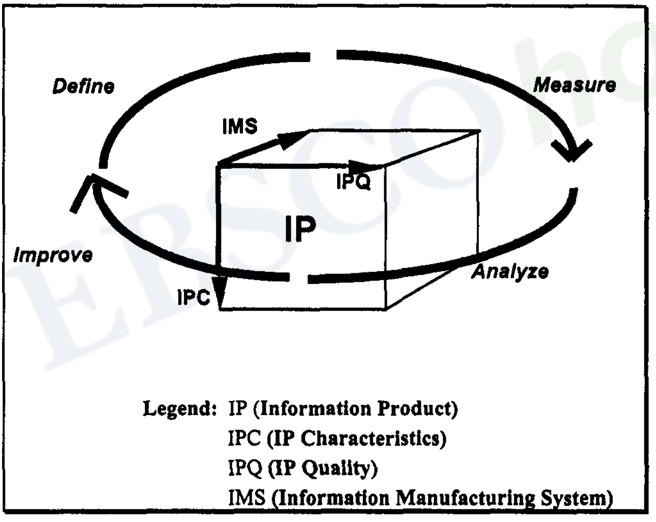

Figure 1.2: A Schematic of the TDQM Framework

(Source: *Communications of the ACM* [32])

In applying the TDQM framework, one must first define the characteristics of the IP, assess the IP's data quality requirements, and identify the ***information manufacturing system*** for the IP [5]. This can be challenging for organizations that are not familiar with the TDQM framework. Our experience has shown, however, that after the previously mentioned tasks have been performed a number of times and the underlying concepts and mechanisms have been understood; it becomes relatively easy to repeat the work. Once these tasks are accomplished, those for measurement, analysis, and improvement can ensue.

Define IP

The characteristics of an IP are defined at two levels. At a higher level, the IP is conceptualized in terms of its functionalities for information consumers just like when we define what constitutes a car, it is useful to first focus on the basic functionalities and leave out advanced capabilities (e.g., optional features for a car such as A/C, stereo, or cruise control).

Continuing with the client account database example, the functionalities are the customer information needed by information consumers to perform the tasks at hand. The characteristics for the client account database include items such as account number and stock transactions. In an iterative manner, the functionalities and consumers of the system are identified. The consumers include brokers, client representatives, financial controllers, accountants, and corporate lawyers (for regulatory compliance). Their perceptions of what constitutes important data quality dimensions need to be captured and reconciled.

At a lower level, one would identify the IP's basic units and components, along with their relationships. Defining what constitutes a basic unit for an IP is critical as it dictates the way the IP is produced, utilized, and managed. In the client account database, a basic unit would be a un-grouped client account.

In practice, it is often necessary to group basic units together (e.g., eggs are packaged and sold by the dozen). A mutual-fund manager would trade stocks on behalf of many clients, necessitating group accounts; top management would want to know how much business the firm has with a client that has subsidiaries in Europe, the Far East, and Australia. Thus, a careful management of the relationship between basic accounts and aggregated accounts, and the corresponding processes that perform the mappings are critical because of their business impacts.

Components of the database and their relationships can be represented as an entity-relationship (ER) model. In the client account database, a client is identified by an account number. Company-stock is identified by the company's *ticker symbol.* When a client makes a trade *(buy/sell),* the *date, quantity* of shares and *trade price* are stored as a record of the transaction. An ER diagram of the client account database is shown in Figure 1.3.

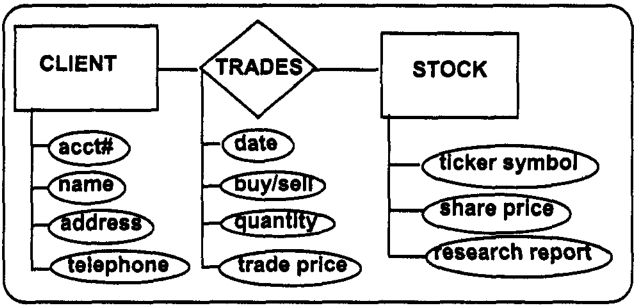

Figure 1.3: A Client Account Schema

With the characteristics of the IP specified, the next step is to identify data quality requirements from the perspectives of IP suppliers, manufacturers, consumers, and managers. We have developed an instrument for data quality assessment and corresponding software to facilitate the data quality assessment task. After data are collected from information collectors, custodians, consumers, and IP managers and entered into the survey database, the data quality assessment software tool performs the query necessary for mapping the item values in the surveys to the underlying data quality dimensions [9]. Figure 1.4 illustrates the capabilities of the software tool.

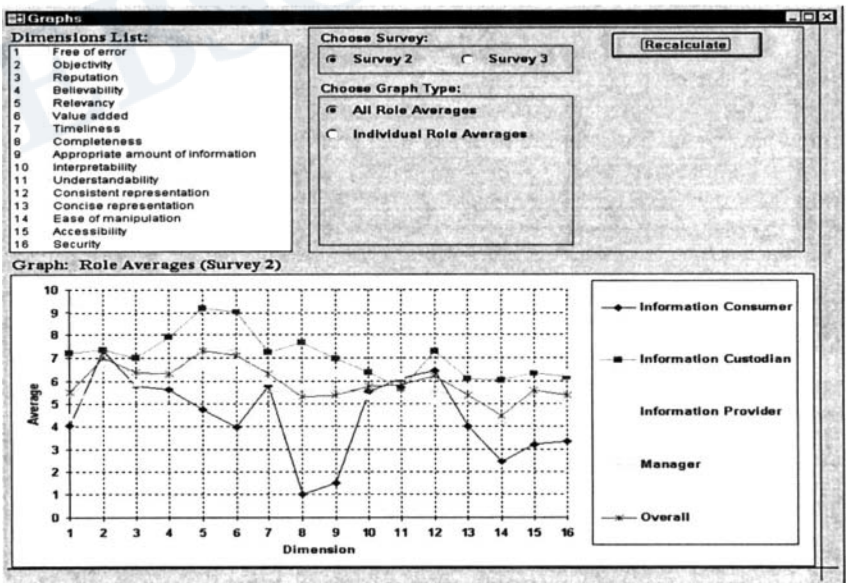

Figure 1.4: Dimensional Assessment of Data Quality Importance Across Roles

As can be seen from Figure 1.4, the result from the first dimension indicates that the custodian believes the IP to be largely free of error (score 7 on a scale from 0 to 10 with 10 being completely free of error), whereas the consumer does not think so (with a score of 4). Both the custodian and the consumer indicate that the IP contains data that are objective and relatively important (score 7). The biggest contrast shows up for Dimension 8, *completeness.* Although the custodian assesses the IP to have complete data (score 7.6), the consumer thinks otherwise (score 1)!

From the IP characteristics and the data quality assessment results, the corresponding logical and physical design of the IP can be developed with the necessary quality attributes incorporated. *Timeliness* and *credibility* are two important data quality dimensions for an IP supporting trading operations. In Figure 1.5, *timeliness* on *share price* indicates that the trader is concerned with how old the data is. A special symbol, "√ inspection" is used to signify inspection requirements such as data verification.

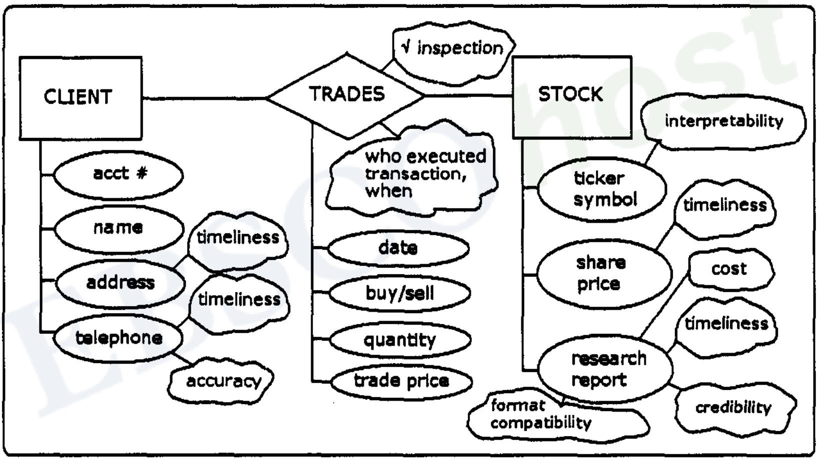

Figure 1.5: Information Quality added to the ER Diagram

(Source: Proceedings - 9th International Conference on Data Engineering [33])

The data quality requirements are further refined into more objective and measurable characteristics. As shown in Figure 1.6, these characteristics are depicted as a dotted-rectangle. For example, *timeliness* is redefined by *age* (of the data), and *credibility* of the research report is redefined by *analyst name.* The quality indicator *collection method,* associated with the *telephone* attribute, is included to illustrate that multiple data collection mechanisms can be used for a given type of data. Values for the collection method may include "over the phone" or "from an existing account".

The quality indicator *media* for *research report* is used to indicate the multiple formats of database-stored documents such as bit-mapped, ASCII or Postscript. The quality indicators derived from "√ inspection" indicate the inspection mechanism desired to maintain data reliability. The specific inspection or control procedures may be identified as part of the application documentation. These procedures might include independent, double entry of important data, front-end rules to enforce domain or update constraints, or manual processes for performing certification of the data.

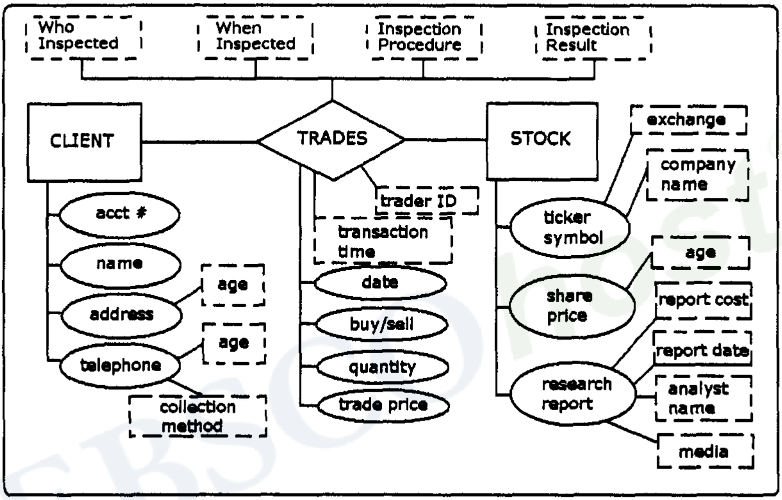

Figure 1.6: A Quality Entity-Relationship Diagram

(Source: Proceedings - 9th International Conference on Data Engineering [33])

Equally important to the task of identifying data quality dimensions is the identification of the information manufacturing system for the IP. Figure 1.7 illustrates an information manufacturing system which has five data units (DU_1 - DU_5) supplied by three vendors (VB_1 - VB_3). Three data units (DU_6, DU_8, DU_{10}) are formed by having passed through one of the three data quality blocks (QB_1 - QB_3). For example, DU_6 represents the impact of QB_1 on DU_2 [5].

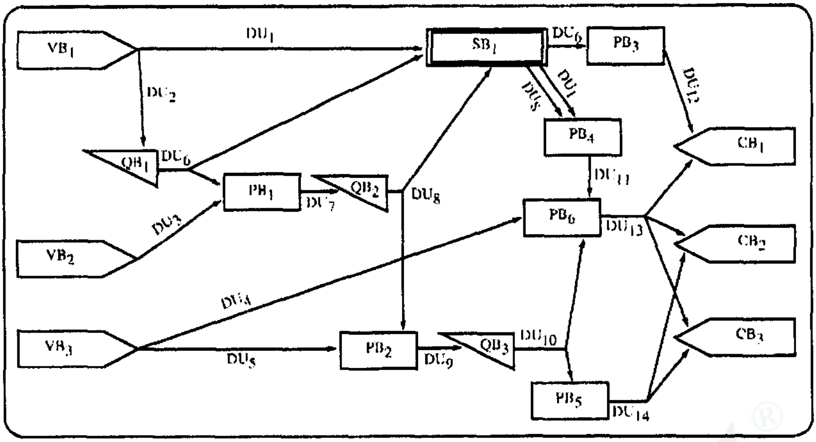

Figure 1.7: An Illustrative Information Manufacturing System

(Source: *Management Science* [5])

There are six processing blocks (PB_1 - PB_6) and, accordingly, six data units (DU_7, DU_9, DU_{11}, DU_{12}, DU_{13}, DU_{14}) that are the output of these processing blocks. One storage block (SB_1) is used both as a pass through block (DU_6 enters SB_1 and is passed on to PB_3) and as the source for data base processing (DU_1 and DU_8 are jointly processed by PB_4). The system has three consumers (CB_1 - CB_3), each receives some subset of the IP. The placement of a quality block following a vendor block (similar to acceptance sampling) indicates that the data supplied by vendors is deficient with regards to data quality.

In the client account database example, identifying such an information manufacturing system would provide the IP team with the basis for assessing the values of data quality dimensions for the IP through the ***Information Manufacturing Analysis Matrix*** and for studying options in analyzing and improving the information manufacturing system [5].

The IP definition phase produces two key results: (1) a quality entity-relationship model that defines the IP and its data quality requirements, and (2) an information manufacturing system that describes how the IP is produced, along with the interactions among information suppliers (vendors), manufacturers, consumers, and IP managers.

With these results from the IP definition phase, an organization has two alternatives. First, the organization can develop a new information manufacturing system for the IP based on these results. The advantage of this approach is that many data quality requirements can be designed into the new information manufacturing system, resulting in quality-information-by-design analogous to that of quality-by-design in product manufacturing. Many of the data quality problems associated with a legacy system can also be corrected with the new system. The disadvantage is that it would require more initial investment and significant organizational change.

Second, the organization can use these results as guidelines for developing mechanisms to remedy the deficiencies of the existing system. However, as the business environment changes, and with it the information needs of the consumers, a new information manufacturing system will ultimately need to be developed.

Measure IP

The key to measurement resides in the development of data quality metics. These data quality metrics can be the basic data quality measures such as data accuracy, timeliness, completeness, and consistency [1]. In our client account database example, data quality metrics may be designed to track, for instance:

- The percentage of incorrect client address zip code found in a randomly selected client account *(inaccuracy)*
- An indicator of when client account data were last updated *(timeliness* or *currency* for database marketing and regulatory purposes)
- The percentage of non-existent accounts or the number of accounts with missing value in the industry-code field *(incompleteness)*
- The number of records that violate referential integrity *(consistency)*

At a more complex level, there are business rules that need to be observed. For example, the total risk exposure of a client should not exceed a certain limit. This exposure needs to be monitored for clients who have many accounts. Conversely, a client who has a very conservative position in one account should be allowed to execute riskier transactions in another account. For these business rules to work, however, the IP team needs to develop a proper account linking method and the associated ontology to make the linkage,

There are also information-manufacturing-oriented data quality metrics. For example, the IP team may want to track

- Which department made most of the updates in the system last week
- How many un-authorized accesses have been attempted *(security)*
- Who collected the raw data for a client account *(credibility)*

Other data quality metrics may measure the distribution of the data quality-related collective knowledge across IP roles. Whatever the nature of the data quality metrics, they are implemented as part of a new information manufacturing system or as add-on utility routines in an existing system. With the data quality metrics, data quality measures can be obtained along various data quality dimensions for analysis.

Analyze IP

From the measurement results, the IP team investigates the root cause for potential data quality problems. The methods and tools for performing this task can be simple or complex. In the client account database example, one can introduce dummy accounts into the information manufacturing system to identify sources that cause poor data quality. Other methods include statistical process control (SPC), pattern recognition, and Pareto chart analysis for poor data quality dimensions over time.

We illustrate other types of analysis through the case of the Medical Command of the Department of Defense, which developed data quality metrics for the information in their Military Treatment Facilities (MTF). In this particular case [8], an analysis of the assumptions and rationale underlying the data quality metrics was conducted. Some of the issues raised were:

- What are the targeted payoffs?
- How do the data quality metrics link to the factors that are critical to the target payoffs
- How representative or comprehensive these data quality metrics are
- Whether these data quality metrics are the right set of metrics

The target payoffs could be two-fold: (1) the delivery of ever-improving value to patients and other stakeholders, contributing to improved health care quality; and (2) improvement of overall organizational effectiveness, use of resources, and capabilities. It would be important to explicitly articulate the scope of these metrics in terms of the categories of payoffs and their linkages to the critical factors.

To provide the best health care for the lowest cost, different types of data are needed, MTF commanders need cost and performance data, Managed Care Support contractors need to measure the quality and cost of their services, and patients need data they can use to know what kind of services they get from different health plans. The types of data needed can fall into several categories: patient, provider, type of care, use rate, outcome, and financial. Based on the targeted payoffs, the critical factors, and the corresponding types of data needed, one can evaluate how representative or comprehensive these data quality metrics are and whether these metrics are the right set of metrics.

Improve IP

Once the analysis phase is complete, the IP improvement phase can start. The IP team needs to identify key areas for improvement such as: (1) aligning information flow and workflow with infrastructure, and (2) realigning the key characteristics of the IP with business needs. As mentioned earlier, the Information Manufacturing Analysis Matrix [5] is designed for the above purposes. Also, Ballou and Tayi [3] have developed a framework for allocating resources for data quality improvement. Specifically, an integer programming model is used to determine which databases should be chosen to maximize data quality improvement given a constrained budget.

SUMMARY

In this chapter, we presented concepts and principles for defining, measuring, analyzing, and improving information products (IP), and briefly described a data quality survey software instrument for data quality assessment. We also introduced the Total Data Quality Management (TDQM) framework and illustrated how it can be applied in practice.

The power of the TDQM framework stems from a cumulative multi-disciplinary research effort and practice in a wide range of organizations. Fundamental to this framework is the premise that organizations must treat information as a product that moves through an information manufacturing system, much like a physical product moves through a manufacturing system, yet realize the distinctive nature of information products. The TDQM framework has been shown to be particularly effective in improving IP in organizations where top management has a strong commitment to a data quality policy. Organizations of the 21st century must harness the full potential of their data in order to gain a competitive advantage and attain their strategic goals. The TDQM framework has been developed as a step towards meeting this challenge.

BOOK ORGANIZATION

The remainder of the book is organized as follows. Chapter 2 presents two seminal research efforts, the Polygen Model and the Attribute-based Model, which capture data quality characteristics by extending the Relational Model. Chapter 3 exhibits a research effort that models data quality requirements by extending the Entity Relationship Model. Chapter 4 introduces a knowledge-based model that provides an overall measure that can be useful in automating the judgment of the quality of data in a complex system that uses data from various sources with varying degrees of quality. Chapter 5 describes an attempt to develop a data quality algebra that can be embedded in an extended Relational Model. These chapters summarize our technically oriented work in the database field within the context of data quality.

Other leading research institutions have also been working on issues related to data quality within the framework of database systems. Chapter 6 profiles the MIT Context Interchange project. The European Union's project on data warehouse quality is presented in Chapter 7. In Chapter 8, we profile the data quality project at Purdue University's database group. Finally, concluding remarks and future directions are presented in Chapter 9.

References

[1] Ballou, D.P. and H.L. Pazer, Modeling Data and Process Quality in Multi-input, Multi-output Information Systems. Management Science, 31(2), 1985, 150-162.

[2] Ballou, D.P. and H.L. Pazer, Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff. Information Systems Research, 6(1), 1995, 51-72.

[3] Ballou, D.P. and G.K. Tayi, Methodology for Allocating Resources for Data Quality Enhancement. Communications of the ACM, 32(3), 1989, 320-329.

[4] Ballou, D.P. and G.K. Tayi, Enhancing Data Quality in Data Warehouse Environment. Communications of the ACM, 42(1), 1999, 73-78.

[5] Ballou, D.P., R.Y. Wang, H. Pazer and G.K. Tayi, Modeling Information Manufacturing Systems to Determine Information Product Quality. Management Science, 44(4), 1998, 462-484.

[6] Bowen, P. (1993). Managing Data Quality in Accounting Information Systems: A Stochastic Clearing System Approach. Unpublished Ph.D. dissertation, University of Tennessee.

[7] CEIS. Commander's Data Quality Assessment Guide for CEIS Beta Test - Region 3.Proceedings of Corporate Executive Information System. 1216 Stanley Road, Fort Sam Houston, Texas 78234-6127, 1996.

[8] Corey, D., L. Cobler, K. Haynes and R. Walker. Data Quality Assurance Activities in the Military Health Services System. Proceedings of the 1996 Conference on Information Quality. Cambridge, MA, pp. 127-153, 1996.

[9] CRG, Information Quality Assessment (IQA) Software Tool. Cambridge Research Group, Cambridge, MA, 1997.

[10] DQJ (1999). Data Quality Journal.

[11] English, L.P., Improving Data Warehouse and Business Information Quality. John Wiley & Sons, New York, NY, 1999.

[12] Huang, K., Y. Lee and R. Wang, Quality Information and Knowledge. Prentice Hall, Upper Saddle River: N.J., 1999.

[13] Huh, Y.U., F.R. Keller, T.C. Redman and A.R. Watkins, Data Quality. Information and Software Technology, 32(8), 1990, 559-565.

[14] Huh, Y.U., R.W. Pautke and T.C. Redman. Data Quality Control. Proceedings of ISQE, Juran Institute 7A1-27, 1992.

[15] ICM (1999). ICM Group, Inc.

[16] Kahn, B.K., D.M. Strong and R.Y. Wang (1999). Information Quality Benchmarks: Product and Service Performance. Communications of the ACM, Forthcoming.

[17] Khang, S.W. (1990). An Accounting Model-Based Approach Toward Integration of Multiple Financial Databases (No.WP # CIS-90-15). Sloan School of Management, MIT.

[18] Laudon, K.C., Data Quality and Due Process in Large Interorganizational Record Systems. Communications of the ACM, 29(1), 1986, 4-11.

[19] Liepins, G.E. and V.R.R. Uppuluri, ed. Data Quality Control: Theory and Pragmatics. D. B. Owen. Vol. 112. 1990, Marcel Dekker, Inc., New York, 360 pages,

[20] Morey, R.C., Estimating and Improving the Quality of Information in the MIS. Communications of the ACM, 25(5), 1982, 337-342.

[21] Nichols, D.R., A Model of auditor's preliminary evaluations of internal control from audit data. 62(1), 1987, 183-190.

[22] Pautke, R.W. and T.C. Redman. Techniques to control and improve quality of data in large databases. Proceedings of Statistics Canada Symposium 90, Canada, pp. 319-333, 1990.

[23] Reddy, M.P. and R.Y. Wang. Estimating Data Accuracy in a Federated Database Environment. Proceedings of 6th International Conference, CISMOD (Also in Lecture Notes in Computer Science). Bombay, India, pp. 115-134, 1995,

[24] Redman, T.C., ed. Data Quality: Management and Technology. 1992, Bantam Books: New York, NY. 308 pages.

[25] Redman, T.C., Improve Data Quality for Competitive Advantage. Sloan Management Review, 36(2), 1995, 99-109.

[26] Redman, T.C., ed. Data Quality for the Information Age. 1996, Artech House: Boston, MA. 303 pages.

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

[28] Tayi, G.K.and D. Ballou, Examining Data Quality. Communications of the ACM, 41(2), 1999, 54-57.

[29] TDQM (1999). Total Data Quality Management Research Program.

[30] TTI (1999). Technology Transfer Institute, Inc. Website.

[31] Wand, Y. and R.Y. Wang, Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM, 39(11), 1996, 86-95.

[32] Wang, R.Y., A Product Perspective on Total Data Quality Management. Communications of the ACM, 41(2), 1998, 58-65.

[33] Wang, R.Y., H.B. Kon and S.E. Madnick. Data Quality Requirements Analysis and Modeling. Proceedings of the 9th International Conference on Data Engineering. Vienna: pp.670-677, 1993.

[34] Wang, R.Y., V.C. Storey and C.P. Firth, A Framework for Analysis of Data Quality Research. IEEE Transactions on Knowledge and Data Engineering, 7(4), 1995, 623-640.

[35] Wang, R.Y. and D.M. Strong, Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12(4), 1996, 5-34.

[36] Yu, S. and J. Neter, A Stochastic Model of the Internal Control System. Journal of Accounting Research, 1(3), 1973, 273-295.