DATA QUALITY AND DATABASE MARKETING

John A. Chopoorian, University of Massachusetts Dartmouth
Omar E. M. Khalil, University of Massachusetts Dartmouth
Mehnaz Ahmed, University of Massachusetts Dartmouth

ABSTRACT

The objective of this study is to elaborate on the role of DBM in marketing strategy, identify data quality barriers, and to discuss ways of improving data quality. The paper is organized accordingly: (i) Marketing and DBM, (ii) Data Quality Issues, (iii) Data Quality Management, and (iv) Conclusions.

INTRODUCTION

Database marketing (DBM) and the allied concepts of relationship marketing and one-on-one marketing are using information technology (IT) to gather and sift through enormous amounts of transactional information and other marketing intelligence to bring the customer into ever sharper focus. Combined with other customer oriented advances such as mass "customized" manufacturing, it is at least theoretically possible to find a customer, for example, for a finely tailored suit, and to deliver the product to the busy executive's office within several days of its order.

This is not far-fetched. To succeed in today's global business environment, companies must be able to provide their customers with value and highly competitive service. Success will require personal knowledge of customers and their needs (e.g., Shermach, 1995 and Hinman, 1996), and to achieve this, companies are increasingly adopting DBM as a strategic tool to drive the marketing activities of a company. Information technology is thus fundamentally changing how a company interacts with and serves its customers.

There is, however, a major impediment for the effective implementation of DBM that is an old nemesis - poor data quality. There are many ways for data in a database to be in error. The quality of customer data may be poor because it does not reflect the real world, or it may not be easily used or understood. "Poor quality customer data can cause immediate economic harm and have a more indirect, subtle effect" (Khalil and Harcar, 1997). Marketing managers, unaware of the quality of the data they use, may assume that the adoption of IT makes it "perfect information." The benefits of DBM can only be fully realized, if sufficient consideration is given to the quality and quality management issues concerning data in databases.

MARKETING AND DBM

The buyer/seller relationship has changed significantly over the past century. Pre-twentieth century store owners built one-on-one relationships with their individual customers and knew and remembered every one of them. In the early 50's marketing made a quantum leap, re-discovering the customer as an individual. Companies oriented to customers and "differentiated" their products to appeal to "market segments" – homogeneous sub-sets of customers with common needs. But with rapid advances in IT and software that can store and interrelate data, one-on-one marketing has been re-born in a new era. Moreover, fierce competition is forcing marketing managers to rethink relationship marketing and to seek ways of building long term relations (e.g., Gronroos, 1990; Shani and Chalasani, 1992; Woods and Remondi, 1996). Thus, the practice of DBM started in the 1980s and continues to rapidly grow in the 1990s (e.g., Brachman, et al., 1996; Strategic Research Corp, 1997).

Companies hold enormous amounts of data about their customers, but typically know very little about them. By using today's sophisticated database technology, DBM turns this raw data into useful, intelligence information. Techniques such as statistical regression analysis, data mining, appending of data overlays and predictive modeling using neural network technology, transform raw data into powerful analytical and decision-making tools that are unique to the company. DBM is based on the premise that not all customers or prospects are alike and that by gathering, maintaining, and analyzing detailed information about customers or prospects, marketers can identify key customer segments within their data files and key prospect segments within acquired outside data, and modify their marketing strategies accordingly. In other words, the entire marketing strategy mix is effected by DBM.

The Process and Data Quality Issues:

While each company will, by necessity, handle things a bit differently, the first and foremost action in DBM is building a database or warehouse of databases by gathering both internal company data and external data about industry, competition and customers. Once databases are established, data in the databases need to be managed. Relevant data is coded with those customer attributes deemed to be important. Coded data is then used to search for trends and interrelations of sales and customer behavior. Effectively managed data gives a company a "memory" that enables it to target the right customers at the right time about the right products (e.g., Foskett, 1997). In other words, DBM is not only communications and promotion related, but helps in all of the "utilities" of marketing.

DBM offers several other benefits that support the marketing effort: (i) it helps spend marketing funds cost effectively, since these funds are concentrated on the company's current customers and very best prospects (e.g., (Cats, 1996; Swanson Russel Associates, 1996); (ii) it improves customer relationships and increases customer loyalty; (iii) it enables marketing managers negotiate more effectively in the budgeting process, since the results from most DBM activities are measurable; and (iv) it serves as a research tool as well as a communications tool.

Although the benefits of DBM are immense, they can only be fully realized, if databases contain high quality data that can be properly analyzed and effectively used. Without quality data, DBM will loose its very function of providing marketers with meaningful patterns and trends. Customer data is never perfect. Errors creep in through inaccurate self-reporting by consumers or during data entry and compilation. Consequently, poor data quality can easily render DBM less effective.

Since data is critical to the DBM process, the quality of such data should be continuously measured and its suitability assessed. Generally, quality is related to conformance to specification and fitness for purpose. Using this view, high-quality data is defined as data that is fit for use by data consumers (Strong, Lee, and Wang, 1997; Wand and Wang, 1996; Armstrong, 1994). It follows that data quality problems are any difficulty encountered along one or more quality dimensions that renders data completely or largely unfit for use (Strong, Lee, and Wang, 1997).

Wang et al. (1993) proposes four quality dimensions for data: (i) accuracy, occurs when the recorded value is in conformity with actual: (ii) timeliness, occurs when the recorded value is not out of date; (iii) completeness, occurs when all values for a certain variable are recorded; and (iv) consistency, occurs when representation of all data value are consistent. Other dimensions of data quality that have been identified include data validation, availability, traceability, and credibility. The quality issues of data thus vary depending on the type of data and the purpose to which it is put.

There is substantial evidence that the quality of data in a company's databases or data files may often be quite poor and a major source of errors and losses (Orman, Storey, and Wang, 1994). For example, more than 60% of surveyed firms (500 medium-size corporations with annual sales of more than $20 million) reported problems with data quality (Kiely, 1992). Also, Hardjono (1993) reports that 60% of information systems (IS) managers have poor data. Yet, relatively little attention is given to the issue of data quality.

In a Wall Street Journal report: "Thanks to computers, huge databases brimming with information are at fingertips, just waiting to be tapped. They can be mined to find sales prospects among existing customers; they can be analyzed to unearth costly corporate habits; they can be manipulated to divine future trends. Just one problem: Those huge databases may be full of junk. ..." (Wand and Wang, 1996).

In a major transportation company, 77% of the reasons for incorrect delivery and, consequently customer dissatisfaction, were related to erroneous data, missing data, mistrusted data, or the inappropriate use of data (McGee, 1992). Additionally, in two commercial databases that contain contact information for companies operating globally, between 40 and 48 percent of the records contained errors in either the address or the telephone number while a further 15 to 28 percent could not be verified (Armstrong, 1994). In constructing its data warehouse, a firm discovered that it had as many as 17 account numbers for one customer and eight different ways of spelling that customer's name (Foskett, 1997).

It is common among subsidiaries within the same firm or among different firms in one country or in a number of countries to have dedicated information systems, each with its own user interface, data model, specialized operations, and storage organization. In this case, while sharing and exchanging data between the participating organizations and coordinating this information is critical (Silberschatz and Zdonik, 1996), the existence of many heterogeneous databases makes sharing data rather problematic.

Furthermore, the data that are created by one group (e.g., sales personnel) in one part of the world could be used for decision making by another group (e.g., marketing managers) thousands of miles away. In other words, data that are gathered for a one business purpose and intended to be stored in a single data base may actually be used for many business purposes and replicated in a number of databases and used by multiple users (Huh et al., 1990). While the initial user(s) of the data may be fully aware of the meaning of the various data items, the other users may not. Consequently, data can be easily misinterpreted.

If asked where data quality problems occur, many marketing managers would probably reply, in data entry. However, analysis reveals that data quality problems may occur at every stage of the "data life cycle", in any part of a business process including that of DBM, and for a variety of reasons. Below are data quality problems categorized by the sources from which they arise - creation, analysis and application - and some potential causes (Firth, 1997):

4. Creation and use of the database itself – causes of quality problems:

   - Database Design: Record and field definitions are too loose, unstructured or not normalized. Schema lacks sufficient validation, and integrity rule
   - Data Aging: The company cannot track the age of data, or has no program to update or enrich data
   - Lack of customer response: Data never fully captured. Customer form is badly designed, or no incentive is given to customer to offer response
   - Fraud: Physical and logical system security is lax or compensating controls are absent
   - Input Error: The system input method is badly designed, or lacks automatic validation

   Human errors easily introduced.

   - Business Rules: System requirements lack adequate or current reference to business rules for data
   - Incorrect Attributing: One attributer may be confused with another. For example a person's age may be confused with their birth date
   - Reference Frame: The data may not be relevant and its measurement carried out inaccurately.

2. The data analysis/mining process- causes of quality problems:

   - System conversions, migrations or reengineering: Inadequate data quality testing on the conversion process. Conversion programs introduce new errors. Reengineering does not consider data context.
   - Heterogeneous system integration: Data is inconsistent or contradictory across systems. Inadequate data quality testing on integration
   - Post integration of heterogeneous systems: Data remains inconsistent or contradictory across systems. Subtleties of poor data quality arise as new scenarios develop
   - Production software: Software requirements relating to the analysis software were incomplete or errors were introduced in the development process. Lack of applied software engineering or production controls
   - Systems internationalization: Overlapping or inconsistent interpretation or usage of codes, symbols, formats due to national differences

3. The application and interpretation – causes of quality problems

   - Policy and Planning: Lack of management attention to data quality management

- Misinterpretation:The marketer may fail to interpret the meaningful patterns and trends correctly. This is because data customers may differ from data producers (Morril, 1996)
- Misapplication: The marketer fails to make correct use of the information by wrongful implementation in terms of business strategies and tactics.

Data Quality Management:

Data quality in a conventional database system has been treated implicitly through functions such as recovery, concurrency, integrity, and security control (Wang et al., 1993). These functions are necessary, but not sufficient, to ensure data quality in the database from the data consumer's perspective. In general, data in a database is normally used by a range of different organizational functions with different perceptions of what constitutes quality data in terms of its quality dimensions (e.g., accuracy, completeness, consistency, and timeliness). Therefore, data quality must be calibrated in a manner that enables consumers to use their own yardsticks to measure the quality of data. Unfortunately, the existing database management systems (DBMS's) lack the capability of explicitly representing the quality of data or allowing consumers to measure such a quality.

In order to improve data quality, two interrelated activities have to be considered (Wand and Wang, 1996). Concern must be given to design, development, testing, maintenance software, and the training of users. In the case of DBM, concern must be given to the implementation of the database and the application of analytical software that will excavate meaningful patterns and trends from that data, apart from storing and retrieving it. And secondly, activity that should be considered by any data quality improvement efforts is the production and distribution of data. This activity involves data originators, data distributors and data consumers, and comes into play when creating the database. This cycle revolves around the sources that generate data, the elements that distribute this data and the users of this data. These two activities are important factors that form barriers to data quality, comprising the areas where most data quality issue arises.

Finally, Wang et al. (1992) propose three general factors that should be considered in the data quality enhancement process:

(i) The definition and measurement of data quality. This is essential in uncovering errors in collecting, storing and distributing data, especially at the database level. Although intuitively understandable, the notions of data quality measurement and clear definitions of data quality are not well defined in current practice.

(ii) An analysis of the economic impact of data quality. Such analysis should address the relationship between high quality data and the successful operation of the business; and alternatively, how low quality data may impact the business. For example, in a transportation company implementing DBM by creating a database of their customers to target sales, poor data quality and usage of the database can become the cause of a large percentage of the delivery misses which in turn can result in a corresponding loss of sales.

(iii) (Continuous) improvement of data quality through both technical and managerial solutions. Solutions can be grouped into three interrelated categories: business redesign, data quality motivation, and use of new information technologies. Business redesign attempts to simplify and streamline the DBM operation to minimize the opportunity for errors. Data quality motivation focuses on how rewards, benefits, training, and perceptions may encourage improved data handling and application by members of the organization. New information technologies give improved procedures for data capture, processing and mining.

CONCLUSION

DBM as a strategic tool can drive the marketing agenda of the organization. It fundamentally changes how the company interacts with and serves customers. Data quality, however, is crucial to the effective implementation of DBM. Without quality data, DBM will lose its very function of providing marketers with meaningful patterns and trends based on timely, relevant and accurate data.

We propose a strategic framework for data quality improvement efforts and suggest that a continuous improvement program should consist of the following essential steps:

- Identify a vision and objectives for data and data quality as part of the strategic planning process and its implementation. Establish a relevant business and marketing information agenda focused on "actionable" data.
- Clearly identify or establish the database marketing function in the organization and the cost-benefit implications of DBM and data quality
- Create and maintain relevant databases using acceptable sources and formats of internal and external data flows.
- Establish organizational responsibility within the DBM staff for data quality, training of personnel and quality improvement.
- Set data quality standards using the dimensions of accuracy, timeliness, completeness and consistency (Wang et al., 1993).
- Select data mining and analysis software that supports the data and data quality vision, the marketing information agenda, or increasingly to support Enterprise Resource Planning (ERP)
- Establish procedures and monitor for effectiveness and data accuracy at common junctures of data creation, analysis and application (Firth, 1997).
- Implement system and data quality (continuous) improvement activities (e.g., data entry, process control, system redesign, etc.) (e.g., Wang et al., 1992; Wang, et al., 1994).

In the information age, the old adage "knowledge is power" needs to be tempered. It is the successful "application" of knowledge that will give it power. And, to achieve this, we need quality data that profitably supports strategic and information needs, or it becomes "garbage in, garbage out." An all too familiar disappointment when computers were first looked upon as a panacea.

REFERENCES

Armstrong, C. (1994). CIQM report on database quality. *Database*, 17(6), 45- 48.

Brachman, R. J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., and Simoudis, E. (1996). "Mining business database." Communications of the ACM. 39(11), pp. 42-48.

Cats, H. 91996). New approach to database marketing scores points in packaged goods. *Direct Marketing*. 59(7), pp. 18-21.

Firth, C. (1997). When do data quality problems occur?. Data quality home page, wysiwyg://main.20/http:11 sunflower.signet.cism.sg1~cfirth/dataquality/dq1.htlm

Foskett, S. (1997). Star wars comes to direct marketing. *Direct Marketing*. 60(1), 62-64.

Gronroos, C. (1990). Relationship approach to marketing in service contexts: The marketing and organizational behavior interface. *Journal of Business Research*. January, pp. 3-11.

Hardjono, H. (1993).A Case Study of the Business Impact of Data Quality in the Airline Industry (Cambridge, MA: MIT Sloan School of Management Working Paper, 1993).

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

Kiely, T. (1992). Keeping Your Data Honest. *CIS*, October 15, pp. 42-46.

Khalil, O., and Harcar, T. (1997). Relationship marketing and data quality management. *The Proceedings of the Society for Advancement of Management International Management Conference*.

McGee, A. M. (1992). Total data Quality Management: Zero Defects Data Capture (Cambridge, MA: MIT Sloan School of Management Working Paper, 1992).

Morrill, J. M. (1996). It's the data stupid!. *Journal of Systems Management*. 47(4), p. 41.

Orman, L., Storey, V., and Wang, R. (1994). Systems Approaches to Improving Data Quality (Cambridge, MA: MIT Sloan School of Management Working Paper).

Shani, D. and Chalasani, S. (1992). Exploring niches using relationship marketing. *Journal of Service Marketing*. Fall, pp. 43-52.

Shermach, K. (1995). Large and small retailers see value in data-base marketing. *Marketing News*, 29(20), p. 8.

Silberschatz, A., and Zdonik, S. (1996). Strategic directions in database systems-breaking out of the box. *ACM Computing Surveys*. 28(4), 764-778.

Strategic Research Corp. (1997). database demographic report.http://www.sresearch.com/search/105444.html

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

Swanson Russel Associates (1997). What is database marketing?. http://www.sramarketing.com/sra/tour/database/what is database.html

Wand, Y., and Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*. 39(11), 86-95.

Wang, R. Y., Reddly, and Gupta A. (1993). An object-oriented implementation of quality data products. http:/web.mit.edu/tdqm/papers/93/pass 2/93-14.html

Wang, Storey, and Firth, C. (1994). A framework for analysis of data quality research. http://web.mit.edu/tdqm/www/papers/93/pass2/93_11.html

Wang, R. Y., Kon, H. B., and Henry, B. (1992). Toward data quality management. http:/web.mit.edu/tdqm/papers/92/92-02.html.

Woods, T., and Remondi, J. (1996). "Relationships vital for high-tech marketers." *Marketing News*. 30(11), p.8.