Customer-Centric Information Quality Management

May 24, 2004

Contributors:
Dr. John Talburt, Acxiom Corporation
Dr. Richard Wang, Massachusetts Institute of Technology
Mark Evans, Acxiom Corporation
Dr. Nalaka Edirisinghe, Massachusetts Institute of Technology
Raïssa Katz-Haas, Cambridge Research Group
Traci Campbell, Acxiom Corporation
Emily Kuo, Cambridge Research Group
Dr. Yang Lee, Northeastern University

Acknowledgements: Research conducted herein has been supported, in part, by Acxiom Corporation, MIT Information Quality (MITIQ) Management Program, and Cambridge Research Group.

HTTP://MITIQ.MIT.EDU/MITIQ/CCIQM.HTM

1. INTRODUCTION

Many businesses are working to optimize their relationships with their customers using customer relationship management (CRM). However, Gartner Group reports "*through 2006, more than 50 percent of all CRM implementations will be viewed as failures from a customer's point of view*."[1] Gartner Group also states that one way to prevent such failures is for organizations to adopt a more customer-centric strategy across the enterprise.

Being "customer-centric" means knowing your customer. A customer-centric approach requires companies to interact consistently with customers across all channels, gather customer data, and gain detailed insight into customer behavior and preferences. To gain this type of insight, businesses need the highest quality customer information. Business strategies for customer-centricity continually evolve and change; information management plans must accommodate and even facilitate these changes so that the quality of customer information remains high enough to support a customer-centric business model. This reality requires re-thinking enterprise customer data management strategies.

This paper introduces an approach to customer information management that significantly increases the quality of customer data. This approach is called Customer-Centric Information Quality Management (CCIQM). This paper discusses the importance of CCIQM for today's corporations and explores some challenges CCIQM addresses.

[1] http://www.businesswire.com/webbox/bw.091001/212532098.htm

2. WHAT IS CCIQM?

CCIQM is an approach to managing customer relationships in a truly "customer-centered" way. This management approach employs methods that significantly increase the quality of customer information to a high enough level required by a customer-centric business model.

Understanding what it means to be "customer-centric" means understanding this business model and the strategies that involves. Most organizations deal with four types of customers:

- Individual
- Household of Individuals
- Business
- Corporate Household (Related Businesses)

The individual customer, sometimes referred to as a *consumer*, is a person with a unique identity. Information about an individual often includes a name, address, phone number, and occupancy status. A household is usually defined as a group of individuals who share the same address, e.g., a family. A similar situation exists with business customers, either an individual business operating at one location, or a corporate household of different-but-related units such as departments, divisions, and branches.

CCIQM addresses problems related to infrastructures where customer contact information is organized in "clusters"; that is, all information related to a customer is linked or grouped together. Information about all types of customers normally exists in multiple data stores, each with its own customer recognition process. Because of this 'scattered' nature of customer data, organizations encounter problems when attempting to adopt customer-centric strategies.

Information quality must be high enough for organizations to balance their ability to:

1. Discriminate between customers
2. Retain and integrate all data related to a single customer.

Customer recognition occurs when customer data integration (CDI) methods and technologies are used to determine the identity of a customer, regardless of how, when, or where that customer interacts with the company; customer recognition *systems* are automated systems for determining the identity of a customer through the information given during a customer transaction and for performing customer synchronization across an enterprise. CCIQM addresses issues encountered in these customer recognition systems used for CRM applications.

3. WHY IS CCIQM IMPORTANT?

To stay competitive today, businesses allow their customers to reach them through multiple touch points, such as web sites, call centers, and sales teams[2]. At the same time, today's customers demand that businesses 'know' who they are across all lines of business, accounts, products, and interactions.

By employing multiple touch points, companies tend to spread out a single customer's information across multiple data stores or silos in different divisions. However, to remain competitive, companies must have a single view of each customer across all touch points. Companies must know their customer well enough to customize its products and services to meet the needs of that customer.

To support this customer-centric business strategy, organizations will need to change their business, information management, and technical strategies from "product- or account-centric" to "customer-centric." Customer-centricity can only be as good as the quality of the customer information upon which it is based. Consider the following example:

> *One of the world's largest and most technologically savvy firms, ABC Company, undertakes a project to build a customer marketing database. The firm employs a staff of over fifty technicians, purchases a leading CDI software package, consults with expert system designers, and "successfully" implements the first release of the database. Once implemented, they discover records containing data about more than one customer. The records have been 'over-matched.' The problem must be corrected without taking another four months to reload records into the database. After repeated consultation with the CDI software vendor, the firm contacts a leading data quality firm for consultation. It takes the data quality firm a half-day to ascertain that the problem can only be corrected by redesigning the database. After spending millions of dollars on the project, management loses faith and cancels the project.*

What happened? Records were 'cleansed' and matched as well as could be expected. But the database design assumed that once records were matched, they were correct, and any 'leftover' name and address records were discarded. Therein lies the problem--once this data is discarded, it cannot be recovered or used to correct errors. Reloading the database each time an error is discovered is not a viable option.

The above scenario reveals how customer-centric solutions can fail to meet business needs. The most challenging aspect of these problems is that they usually are not found until 24 to 36 months after solutions are up and running. Then, it is quickly discovered that the problems cannot be corrected without redesigning the system. Errors continue to go uncorrected, and the solution loses both credibility and usage until it is (hopefully) replaced.

---

[2] http://www.destinationcrm.com/articles/default.asp?ArticleID=3947

The following are the types of challenges that CCIQM addresses[3]:

1. How do you handle multiple customer data stores and their processes?

Most organizations moving towards enterprise customer-centricity already have multiple, customer data stores, each supporting its own customer recognition process. Often, the quality of information falls short of the levels needed for applications.

2. How do you benchmark the quality of customer-centric data stores?

Customer information is constantly evolving. Changes occur for many reasons-- life stage changes, data corrections, changed business practices, and new information. A consumer may use several different names and addresses over his/her lifetime.

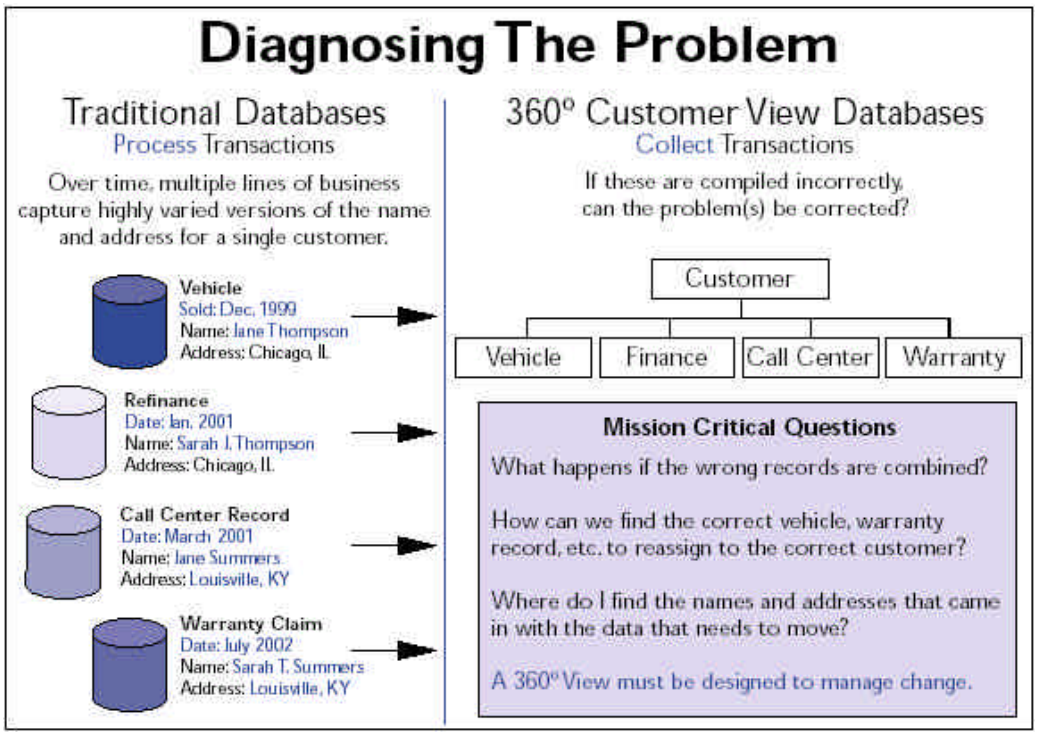

Figure 1 *Diagnosing the problem*

The customer in Figure 1 above uses a different name or address each time she conducts business with the company over a two-year period. The company receives this information from multiple sources such as a call center record and a warrantee claim. When this information is compiled, only the most recent name and address information is retained. Information about all business relationships is tied directly to the final name and address record, and all other data are discarded.

[3] Successful Customer Recognition: Delivering consistent quality in customer-centric databases. An Acxiom White Paper, January 2004.

Multi-sourced, customer data stores cannot achieve 100% accuracy. When dealing with customers, any percentage of errors – no matter how small – must be dealt with. More importantly, 'clean' data is not necessarily *accurate.* Consider the following examples:

Cleansed Data	Accurate Data	Problem
John Smith	**John Smith Jr.**	The missing suffix can cause the original record to be matched to either John Smith Sr. or left as a new person when it should have been integrated with John Smith Jr.
John Smith **113 Oak Street**	**John Smith** **131 Oak Street**	113 Oak Street is a valid address. It passes all address hygiene standards. However, John Smith actually lives at 131 Oak Street. These two records will remain separate and be treated as two different clients.
John Smith **131 Oak Street, Apt 10** **Kathy Smith** **131 Oak Street, Apt 10**	**John Smith** **131 Oak Street, Apt 210** **Kathy Smith,** **131 Oak Street, Apt 10**	The apartment number could cause John to be incorrectly associated with Kathy. Once this is discovered, any relationships attributed at this level (such as joint vehicle ownership) may be difficult or impossible to separate.

Figure 2 *Cleansed data vs. accurate data*

Looking at the information in Figure 2, it is impossible to know which name and address was used to compile the Vehicle, Loan, Call Center, or Warranty information. In addition, it would be useful to develop a scorecard to identify and measure error rates. Thus, customer-centric solutions require proactive quality management that ensures accurate information at the completion of each update.

Given the variability of information and likelihood of errors, several questions come to mind:

- How will errors be handled?
- Can the correct customer information be restored?
- How can the effectiveness of recognition software be measured to ensure delivery of error-free customer data?
- Is it possible to incorporate a self-healing mechanism whereby as more complete and accurate information is received, groups of records that were incorrectly combined based on older information can be separated and re-combined correctly based on new information?
- How do you measure the degree of self-healing?

3. How do you handle left over information?

Direct marketing systems consolidate information for the purpose of managing segments of customers rather than individuals. Therefore, accuracy requirements are not as strict as those needed for one-to-one interactions, which often occur in real time. As a result, marketing database designs include a number of practices that cause serious problems for customer-centric solutions. Data 'cannibalization', or the discarding of left over information, is one such example. As evidenced in the previous ABC Company scenario, customer-centric databases require expansive, multi-faceted, historical data management infrastructures to ensure that incorrectly matched records can be disassembled and re-applied when errors are discovered.

4. How do you handle customer clustering and discrimination?

Clustering and discrimination precision are two key areas surrounding inter-record and knowledge base structures, particularly those that enable customer recognition. The following research questions need to be addressed when handling this challenge:

- How can the need to discriminate between customers be balanced with the need to cluster all data related to a single customer?
- What is the quality (i.e., accuracy) of clustering?

5. How can you identify quality metrics for customer recognition systems?

A customer recognition system is only as good as the quality of its recognition. As such, quality metrics relevant to customer recognition systems need to be defined and validated. The quality of a customer recognition system can be measured along several dimensions, including the accuracy of recognition, timeliness of the recognition request, and timeliness of data recognition.

Some research questions in this area include:

- Is there an objective method to compare the relative accuracy (recognition, clustering, and discrimination) of two or more competing customer recognition systems?
- Is there an objective method to compare whether a recognition engine rule-change results in better or worse quality of recognition?
- Are there methods and techniques that can be used to automatically generate an optimal set of recognition rules given a set of transactions that have already been linked (associated) with the correct customer?

6. How do you maintain the quality of customer recognition systems over time?

While identifying and measuring the quality of customer recognition systems is important, the time dimension is another factor that needs to be addressed. A common problem facing customer recognition systems is the degradation of their accuracy over time. The following research questions need to be addressed in handling this challenge:

- Is it possible to measure and correlate the effect of update transaction accuracy with changes in the overall accuracy of a recognition system?
- Given the average accuracy and rate of update within a recognition system, is there a method for predicting the length of time that the system will retain an acceptable level of accuracy?

4. RESEARCH AGENDA

The CCIQM Work Group at MIT is currently pursuing research in three main areas:

1. **Establishing a similarity index to compare the relative grouping performance between customer recognition systems.**

   The primary function of a customer recognition system is to identify the transactions belonging to a particular customer. So for a given set of transactions, a particular recognition system will group the transactions into disjoint subsets--each subset representing those and only those transactions belonging to one customer. In mathematical terms, these subsets comprise a partition of the set of transactions. Consequently, the overall behavior of a recognition system can be described in terms of the partition it generates for the set of transactions it processes.

   One of the first goals of the CCIQM research is to investigate various methods for measuring the similarity between partitions and to select one that will be most appropriate for the context of customer recognition. A number of such measures have been developed for cluster analysis with applications in a wide variety of areas -- from the analysis of gene expression data to acoustic studies of the seafloor.

   The ability to measure the similarity between two recognition system outcomes is crucial to establishing information quality metrics for two reasons:

   First, an automated and objective measurement of the *consistency* between two systems (or two versions of the same system) could be invaluable in tuning the rules of a recognition engine.

   Secondly, if the correct partition of a set of transactions is known, then the similarity measure can provide an automated and objective way to determine the *accuracy* of a given recognition system acting on the same set of transactions.

   In current practice, these evaluations are often subjective and based on manual inspection of a relatively small number of grouping outcomes.

   Preliminary work is underway to evaluate a binary factoring index developed by Dr. Talburt and Dr. Wang versus the Rand Index[4] and Adjusted Rand Index[5] based on pair-wise clustering counts.

2. **Implementing a CCIQM Test Bed for Customer Data Integration and Customer Recognition.**

   In order to carry out experiments related to the quality measures outlined in this paper, the Work Group proposes establishing a CCIQM Test Bed System based on a highly flexible and customizable recognition engine developed by Acxiom Corporation. The system is ideal for this application because it is completely open (i.e., does not rely on any built-in proprietary logic) and can be readily configured to implement almost any combination of matching and grouping rules.

---

[4] Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, W.M. Rand, 1971.

[5] Comparing partitions. *Journal of Classification*, L. Hubert and P. Arabie, 1985.

3. **Researching data architecture, data governance, and data sharing best practices in the context of CCIQM.**
   Research will be conducted to determine the best practices of data architecture, data governance, and data sharing.

   Data architecture is an element of an enterprise architecture that comprises a data model and assigns accountability for data integrity. It reflects business area entities with attributes and establishes accountability for this information in business process improvements.

   Data governance is the practice of making enterprise-wide decisions regarding an organization's information holdings. Data governance includes the determination of data sources, responsibilities for integrity, defining requirements for business process development and change, and mechanisms of arbitration differences among stakeholders.

   Data sharing is the practice of providing data from one information source to an information consumer in response to a business requirement. Data sharing requires businesses to define specific policies, processes and technologies.

5. CONCLUSION

Successful Customer Relationship Management (CRM) requires customer-centric strategies across all areas of an organization's business-- from its culture to its integration of business processes. These strategies use customer information as a common foundation in providing continuity of service across all business functions, yet allows each service to optimize around its assigned mission. Because customer-centricity can only be as good as the quality of the customer information upon which it is based, it is critical to assure the highest quality, consistency, and flexibility when managing customer information. Customer-centric Information Quality Management recognizes that ensuring customer data quality requires rethinking quality management processes and infrastructures.

For further information about the CCIQM Work Group, please contact either Dr. Richard Wang (rwang@mit.edu) at MIT Information Quality Program or Dr. John Talburt (John.Talburt@acxiom.com) at Acxiom Corporation.