R-14-7181
S. Nelson, R. Singhal, W. Janowski, N. Frey

Strategic Analysis Report
26 November 2001

Customer Data Quality and Integration: The Foundation of Successful CRM

Management Summary

With the goal of creating a more satisfying experience for their customers — and boosting customer loyalty and retention — many enterprises have invested huge sums and effort into deploying customer relationship management (CRM) systems. Unfortunately, many of these same enterprises still fail to understand that a CRM solution is only as good as the quality of the customer data that feeds it.

High-quality, well-integrated customer data is the cornerstone of a successful CRM effort. It is also the key achieving several critical benefits — such as eliminating excess operational costs caused by redundant data, and enhancing revenue through improved customer targeting and retention. Achieving these benefits requires close attention to customer data channel and analysis considerations, and a range of data quality tools and techniques. It also requires that enterprises understand and investigate the benefits of effective customer data integration (CDI).

The backbone of any CRM process, CDI can best be described as the combination of the technology, software, processes and services needed to achieve a single, accurate and complete view of the customer across multiple sources of customer data, databases and business lines. Successful CDI will enable enterprises to recognize customers instantly — and to access relevant customer information dynamically, regardless of the interaction “touchpoint.”

CDI is quickly becoming a buzzword in CRM circles, and enterprises and vendors are starting to talk about it as an established market segment. Many see it as merely data-cleaning revisited, but its components go beyond that. Bringing together the core data functions of data hygiene, linking (i.e., matching records), grouping (i.e., viewing records based on business rules) and customer recognition, CDI can reduce operational and marketing costs and enhance revenue-generating opportunities through increased customer satisfaction and the identification of new customers.

This *Strategic Analysis Report* provides a comprehensive overview of customer data quality considerations and CDI technologies. It begins by examining the nature of customer data and its role in effective CRM; the three basic data types and their interaction in a customer database; and important considerations regarding data channels and analysis for CRM. Next, it explores customer data costs and benefits, and provides case studies of the returns several companies have achieved on their CDI investments. It also examines customer data tools and privacy issues, and provides an overview of the CDI technology market.

It concludes with a summary of the steps enterprises should take to assemble, cleanse and integrate data within the enterprise’s customer data repository, which should serve as the foundation of its CRM efforts.

Although it is rarely a trivial undertaking, developing and maintaining a high-quality, integrated customer data repository is well worth the effort. It may not drive value in and of itself, but it is the means to achieve the cost reduction and revenue enhancement benefits described in this *Strategic Analysis Report.*

CONTENTS

FIGURES

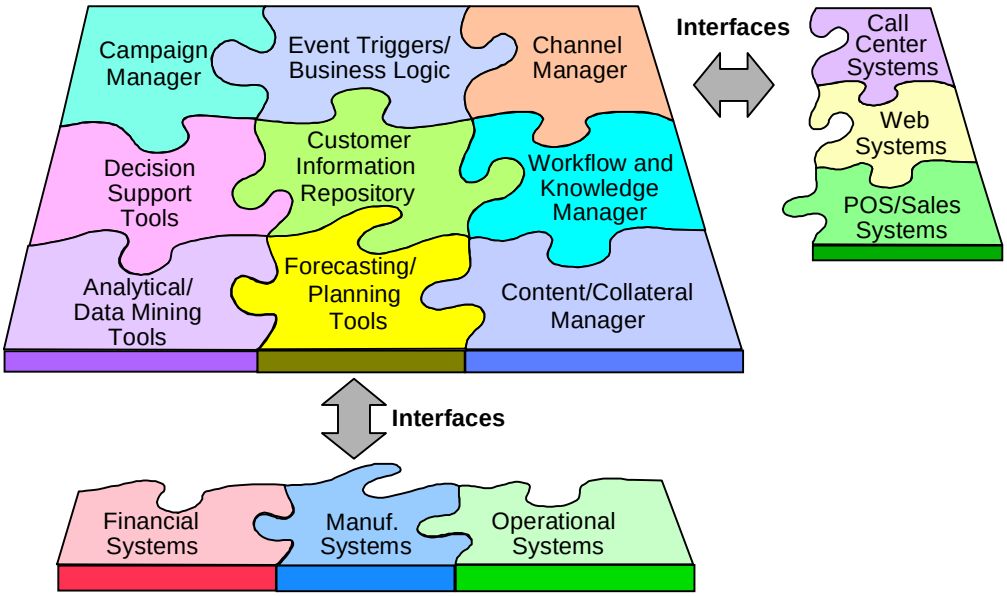
1.0 The Role of Customer Data in CRM

The term CRM — encompassing the largest grouping of IT concepts to date — refers to the overall concept of moving ownership of the customer away from the individual departments and consolidating it at the enterprise level. Under the CRM concept, individual departments are still responsible for customer interactions, but the enterprise is responsible for the customer. Enterprises approach CRM by automating each customer touchpoint: Sales force automation, the Internet, point of sale (POS) technologies and call centers are all pieces of CRM, although they are not substitutes for it.

The backbone of any CRM solution is the data behind it. Enterprises have invested considerable effort and expense in deploying CRM solutions to make their customers' experiences more satisfying, increase retention and foster loyalty. Despite this expense and effort, however, many of these enterprises still fail to understand that a CRM solution is only as good as the quality of the customer data that feeds it.

A critical success factor for CRM is deploying technology that provides multiple levels of data access between all CRM-related applications. Data is important to CRM from a technology perspective and as a whole. Figure 1 illustrates how the pieces of a CRM solution interact to:

- Capture customer data
- Consolidate it in a central database for analysis
- Distribute the resulting information
- Use this information through a variety of communication media

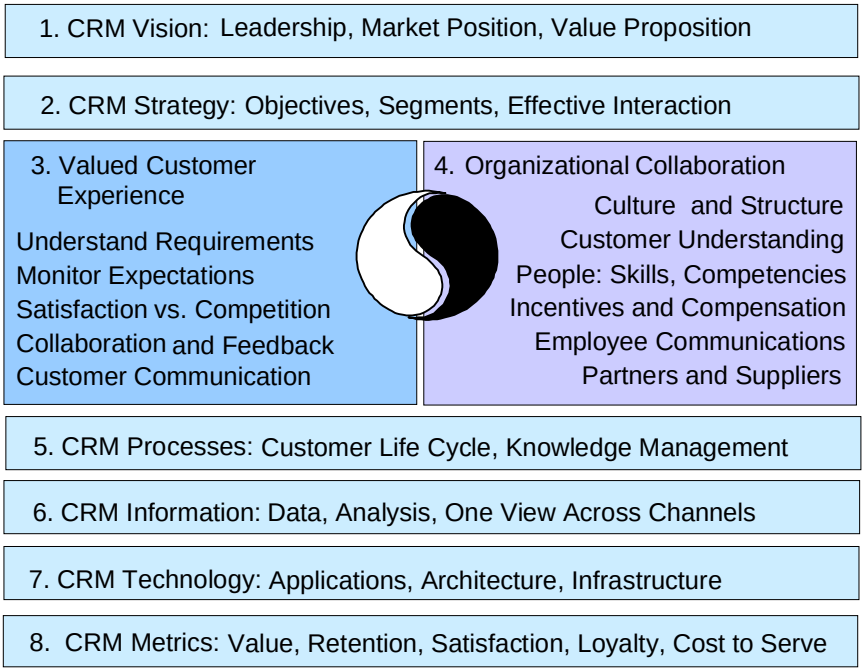


Source: Gartner Research

Figure 1. Architectural Components of a CRM Solution

An effective CDI solution is a vital component of CRM. CDI can best be described as the combination of technology, software, processes and services required to achieve a single, accurate and complete view of the customer across multiple sources of customer data (internal and external), databases and business lines. Successful CDI will enable enterprises to recognize customers instantly — and access their relevant information dynamically, regardless of the interaction touchpoint. CDI is really a data management process that distributes customer data to points of interaction, in a timely and accurate manner.

The "single view" of the customer that CDI provides — together with the effective analysis and maintenance of customer data — is at core of the sixth of Gartner's "Eight Building Blocks of CRM" (see Figure 2 and *Research Note* DF-14-2111). This "CRM Information" building block is the focus of this *Strategic Analysis Report.*



Source: Gartner Research

Figure 2. The Eight Building Blocks of CRM

2.0 Types of Customer Data

Enterprises must have three types of data to effectively manage their customer relationships:

- Descriptive data
- Relationship data
- Contextual data

To appreciate the effectiveness of a CRM solution, one must understand how these information types interact to provide a coherent picture of the customer relationship.

2.1 Descriptive Data

Descriptive data focuses on the customer, which could be an individual, a household, a business or some combination of the three. Demographic, lifestyle and psychographic data fit into this category, because all of these types of information attempt to describe some aspect of the customer. Because it is customer-focused, this information is common across multiple industries and relationships, although some elements will be more important in certain industries than others.

Much of this data comes from operational systems within the enterprise or from external data providers. This information is easily available to most competitors within an industry and, therefore, yields little competitive advantage.

2.2 Relationship Data

This data set includes details on the transactions and interactions that comprise the relationship between the enterprise and its customers. Acquiring relationship data has proven to be the biggest challenge for many enterprises, because they must strike a balance between collecting too much and too little data. Conceivably, the enterprise could collect vast amounts of data about every interaction, but in reality, only so much can be assimilated. Therefore, choices must be made about which interactions, and which data about these interactions, are most valuable.

Because this data personalizes the customer/enterprise relationship, it overcomes the lack of differentiation characteristic of the generic customer data that most enterprises already have. It can also pose challenges in the areas of systems integration and data management (e.g., collecting and managing data from multiple channels and systems before identifying which data points are actually valuable).

2.3 Contextual Data

Contextual data is the least common type of information for an enterprise to have; however, without the information that this data can provide — an understanding of the context of the enterprise's relationship with its customers — the enterprise is unlikely to maintain a strong customer relationship as circumstances change.

Consider, for example, a bank that offers a fixed-rate mortgage. Customer loyalty will likely vary inversely with interest rates, so little data about the customer's description — or his or her previous relationship with the bank — will be as important to his or her loyalty as the context of rising or falling interest rates.

Because contextual data is both diverse and unstructured, it is difficult to integrate with operational customer relationship systems. Identifying the relevant pieces of information will be the first major challenge for the enterprise. Because this information describes the context of the enterprise/customer relationship, what is relevant will vary considerably among industries, and even among enterprises within an industry. For some enterprises, weather forecasts may be critical; for others, government deregulation, competitor behavior, current-affairs television programs or changes in financial markets may be important factors.

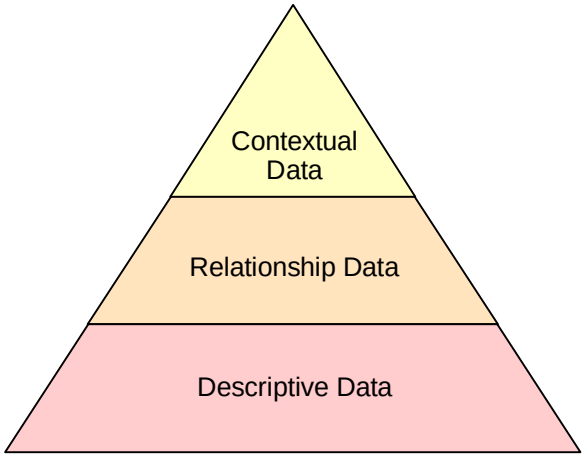
2.4 Data Types and the Marketing Transaction Database

These three data types form the foundation of the data that will go into the marketing transaction database:

- *Descriptive data* occupies the bottom of the customer data "pyramid" (see Figure 3). This data, however, must be the most accurate, because it feeds into the relationship data at the next level.
- The *relationship data*, in turn, will be used by the enterprise to understand its relationships with customers. The key to gaining this understanding is to identify which customer relationships are the critical ones, so that the enterprise can segregate these key relationships from lower-value ones and pay special attention to nurturing them.
- *Contextual data* is probably the hardest to obtain, but is critical for enterprises to have.

Ideally, descriptive data should feed the relationship data, which in turn should feed the contextual data — but this is not always the case. For example, some enterprises do not gather much descriptive data, but have considerable data of both the contextual and relationship types. In these cases, Gartner

recommends that enterprises obtain the descriptive data from data providers, and substantiate it by comparison against other third-party data.



Source: Gartner Research

Figure 3. The Data Pyramid for a Marketing Transaction Database

3.0 Channel Considerations

Traditionally, customer information has flowed in one direction — from operational systems to the customer data repository or data warehouse — and has supported customer-oriented applications in the back office. Today, however, the information flow is increasingly bidirectional: Customer information not only flows from operational systems to the customer data repository, but it also travels from the data store back out to the operational systems and points of customer contact.

Information usage is spreading beyond back-office analysis and applications, opening access by casual information users for front-line customer support applications. Functions supportable at the touchpoint include the delivery of customized messages, branch cross-selling prompts, fee waiver and rate discretion support, and call center cross-selling prompts. Figure 4 describes some of the channels that can be used to deliver customer information.

For several years, enterprises have used data-mining applications to uncover hidden patterns in their customer and transaction data. The advent of the Web has provided enterprises with perhaps the largest single source of customer information ever — and also probably the single biggest source of corrupt data.

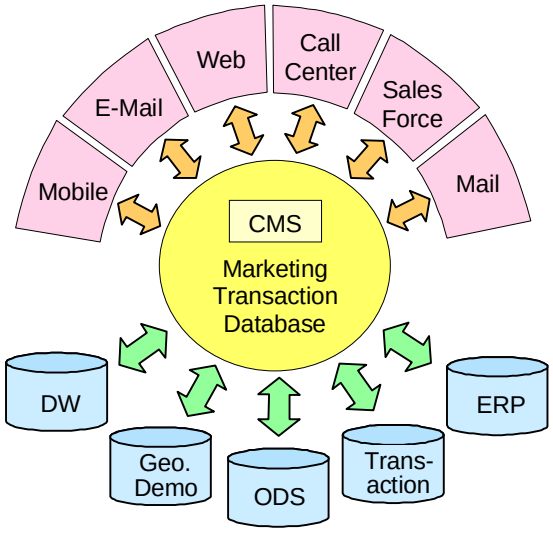
Organizations today handle customer interactions emanating from multiple sources, including e-mail, mobile, Web, fax and telephone channels. As enterprises seek to leverage more current — even real-time — data, new data-mining approaches are necessary. An emerging need for enterprises is to incorporate real-time data into decision-making models. These models will in turn trigger optimized, automated responses, or provide the end user with suggested courses of action during normal operations.

The value of this approach is that enterprises can react quickly and efficiently to problems and opportunities. For this approach to be successful, enterprises — particularly their marketing departments — must start thinking differently in two areas of CRM:

- First, they must stop thinking in terms of individual channels (e.g., direct mail or the Web) when communicating with customers. The customer views the enterprise as a single entity with multiple contact points. To effectively serve its customers' needs, the enterprise must structure itself as the customers perceive it, and present them with a unified set of offers and a high-quality service

experience, regardless of channel. This is the concept of time/place/process convenience, which enables customers to interact with enterprises when they want, how they want and where they want.

- Second, they must discard the notion that marketing cannot be automated. The marketing function must be automated, although doing so is a very difficult process. The permutations of even a handful of channels mixed in with a few different offers quickly become too complicated for any company to manage. As such, enterprises must think in terms of what it will take to automate their marketing function, and how this automation will benefit both the enterprise and its customers.



CMS Campaign management system
DW Data warehouse
ERP Enterprise resource planning
ODS Operational data store

Source: Gartner Research

Figure 4. Multiple-Touchpoint Architecture

4.0 Customer Data Analysis for CRM

How does customer data fit into overall picture of CRM? Building an understanding of the customer relationship requires different types of analysis of customer data, depending on how an enterprise will eventually use this data. This, in turn, requires an understanding of the difference between operational and analytical CRM:

- **Operational CRM** is the business strategy that focuses on the day-to-day management of the customer relationship across all points of customer contact, enabled by sales and service technologies.
- **Analytical CRM** is the component of the CRM business strategy that drives increased customer intelligence and makes information actionable across all touchpoints. Key components of analytical CRM include:
  - *Customer segmentation analysis* to develop formal segments, or ad hoc selections customers for one-time-only marketing campaigns
  - *Customer profitability analysis* to derive profitability information and predict lifetime value, enabling more profitable marketing, sales and service initiatives

- *Predictive modeling* to forecast likely product purchases, customers most likely to buy and customers most likely to be lost by the enterprise
- *"What if" analyses* to determine how specific marketing, sales and service strategies will impact customer profitability
- *Real-time event monitoring and triggering,* which leverages customer insight to take advantage of opportunities as they unfold
- *Campaign management (or relationship optimization),* which uses analytics to develop, execute, manage and measure marketing campaigns across multiple touchpoints
- *Personalization* across all points of customer contact to deliver the most relevant message to a customer, at the right time, via the most appropriate channel

A successful CRM strategy dictates that both analytical and operational CRM be integrated, and that enterprises understand the difference between segmenting customers (analytical CRM) and targeting them (operational CRM). Segmentation is about understanding behaviors and characteristics common to groups of customers. Many enterprises think this is synonymous with customer targeting, but it actually serves a different purpose.

Segmentation enables enterprises to understand the dynamics of customer relationships and is used in two key ways:

- Determining characteristics of profitable customer segments, enabling enterprises to target prospect groups more effectively and aid in the process of targeting current customers
- Identifying data elements that indicate a change in the relationship, enabling enterprises to prepare for future events affecting other customer groups

5.0 Customer Data Costs and Benefits

Costs and benefits are associated with any asset, and data is no exception. The cost of storing data is comparable to that of any other asset under the universal life cycle. Like other resources, data has only a *potential* value because if it is bad, or not used, it is a simply a cost center.

When it is of high quality and used effectively, however, customer data holds the key to achieving numerous benefits — including "hard" benefits (e.g., in terms of cost savings), which can be readily quantified, and "soft" benefits (e.g., in terms of revenue enhancement), which are more difficult to measure but are no less important to the value of customer data.

In this section, we explore the costs and benefits associated with customer data and CDI.

5.1 Customer Data Costs

Before exploring the benefits of having good quality data, enterprises must consider the expense of actually having data to begin with. These costs can be separated into four categories:

- Planning and data acquisition
- Maintenance of data
- Deletion or disposal of data
- Applying data

5.1.1 Planning and Data Acquisition

Before an enterprise starts storing data, it needs a mission statement with explicit objectives and strategies for how it will use the information. In the process of understanding, planning for and acquiring data, enterprises will incur expenses for such activities as modeling and human-capital costs in terms of hours spent on projects. An important data acquisition cost consideration is the IT infrastructure cost (i.e., the cost of hardware, software and support) associated with storing data.

5.1.2 Maintenance of Data

Data maintenance is another cost center for an organization. After data has been acquired and stored in databases, enterprises must maintain and update it continually. Otherwise, it will become stale and useless — a phenomenon known as “information decay.”

Data maintenance costs are often compounded by the too-common tendency for enterprises to have separate, duplicate sets of data throughout the enterprise. The extra cost associated with having duplicate data in two or more databases is known as “redundancy cost.”

5.1.3 Deletion or Disposal of Data

When the data is stale or no longer needed, enterprises should delete it. Discarding unneeded data can be a benefit in itself, because it frees resources and creates value by leaving only good data on which decision-makers can base strategic direction, and customer care representatives can service customers.

5.1.4 Applying Data

An enterprise can follow the preceding steps faithfully but still not derive the full benefits of high-quality data; these processes yield data with a “potential” or “latent” value, but improper application can render the best-quality data useless. Proper, productive data usage requires training of staff, allocating financial and material resources, and finding third-party resources against which to benchmark contextual data to determine its value.

5.1.5 The Cost of Poor Data Quality

High costs are associated with owning data, and Gartner believes that having bad data can increase these costs by a factor of 10 when one considers the costs arising from bad business decisions and poor CRM based on such data. Enterprises should also investigate the underlying reasons behind their poor-quality data; in doing so they will see the overall impact of high-quality data on their business decisions and customer service.

Bad data can creep into an enterprise’s database through several avenues, including:

- Mergers and acquisitions
- Disparate data sources
- Multiple platforms
- Disparate architectures

5.2 Customer Data Benefits

The pitfalls of poor data quality provide the context to fully appreciate the benefits of high-quality data. Although far from a trivial task, developing and maintaining a repository of high-quality customer data is

worth the effort. It may not drive value in and of itself, but it is the means to achieve several important benefits, which include:

- Cost savings from the removal of redundant customer data
- Increased revenue from identifying and targeting first-time customers
- Enhanced revenue from higher customer satisfaction and retention
- Savings in operational costs

5.2.1 Cost Savings From the Removal of Redundant Customer Data

In any Fortune 500 company, the biggest problem in managing customer data is consolidating all databases and keeping only unique entries. In interviews with Gartner, the top five data integration vendors report that the one consistent problem plaguing their clients is the difficulty of determining their true number of unique customers.

This inability to know how many unique customers an enterprise has — not just the number of customer name entries — leads to higher costs (e.g., excess postage and printing costs due to duplicate mailings of marketing materials). Therefore, a key benefit to high-quality data is the savings generated by removing this redundancy.

5.2.2 Increased Revenue From Identifying and Targeting First-Time Customers

The loss of new customers is another cost associated with fragmented or dispersed databases. In 2000, more than 58 million U.S. consumers engaged in online transactions generating $28.5 billion in sales, according to the U.S. Department of Commerce. This number will continue to grow, and enterprises cannot afford to miss the opportunity to service these new customers.

By integrating their customer databases so that they can identify new customers (e.g., first-time user entries in their databases), enterprise can use that information to direct the marketing efforts toward a particular individual to develop the customer relationship. This results in increased revenue through new sales opportunities that would otherwise be missed.

5.2.3 Enhanced Revenue From Higher Customer Satisfaction and Retention

It is often noted that attracting a new customer is 10 times as expensive as retaining an existing one. And with the Internet, enterprises are just “one click away” from losing a customer if he or she is dissatisfied with any element of his or her business transaction.

By developing an accurate and consolidated profile of the customer, enterprises can use that knowledge to tailor their interactions specifically to that customer. An accurate picture of the customer enables more customized, targeted interactions. As the customer’s relationship with the enterprise evolves, it becomes increasingly difficult for a customer to establish the same level of intimacy with the enterprise’s competitors. Through this enhanced customer satisfaction — and enhanced loyalty through the intimacy of their interactions — enterprises can retain more customers and enhance revenue.

5.2.4 Savings in Operational Costs

High-quality data yields a number of operational efficiencies and savings. The internal operational benefits that can be achieved from ongoing data quality review, and the resulting deletion or disposal of redundant data, include:

- Reduced redundancy costs (i.e., the cost of storing the same data in two or more databases)
- Reduced infrastructure costs (the cost of the infrastructure associated with the storage of data)

Enterprises must recognize that data cleaning is an ongoing effort rather than a one-time activity, and should place the ability to perform ongoing data quality review on their list of "must haves." This maintenance must include efforts to maintain levels of data cleanliness and integrity to prevent data degradation. Although it may not be the front-facing functional module that business units ask for, it's a necessity to achieve ongoing cost savings.

5.3 CDI: Hard and Soft Benefits and ROI Case Studies

Gartner believes that enterprises have done little research on calculating their real return on investment (ROI) for implementing CDI solutions because so many items that make up ROI calculations are "soft" metrics. For example, the classic RFM (recency, frequency, monetary) model calculates customer "quality" ratings based factors such as customer retention, customer acquisition and promotional features. These elements contribute to the opportunity cost for an organization, but are difficult to calculate and assign a "hard" number to. Similarly, improving data quality can lead to intangible cost savings by increasing the operational efficiency of databases, and by improving the transactional efficiencies of related processes.

This said, many organizations are also guilty of failing to understand the importance of data quality, or the cost savings that can be achieved from having a well-integrated database. In Figure 5, Gartner has listed some of the areas in which "hard" benefits (in terms of cost savings) and "soft" benefits (in terms of revenue enhancement) can be obtained from CDI.

The items shown in Figure 5 can be used to construct a cost model for calculating the true savings associated with the use of CDI. The following case studies provide examples of some of the hard and soft benefits an enterprise can reasonably expect.

Hard Benefits — Cost Savings

- Data collection and integration savings
- Project cost savings
- Data quality management personnel savings
- Database and data warehouse processing savings
- Faster call handling on inbound inquiries
- Prospect solicitation savings
- List verification/net names savings
- Prospect derogation suppression savings (financial services)
- Customer direct-mail savings
- De-duping mailing files
- Returned mail
- Customer privacy suppression/ opt-out marketing cost savings
- Customer privacy mail notification savings

Soft Benefits — Revenue Enhancement

Cross-Selling

- Revenue enhancement from improved modeling, better targeting and higher conversion rates
- Revenue enhancement from real-time, cross-database event detection
- Revenue enhancement from real-time, customer-facing cross-selling
- Revenue enhancement through finding additional customers to cross-sell

Up-Selling

- Revenue enhancement from better modeling, better targeting and higher conversion rates for up-selling
- Revenue enhancement from real-time, customer-facing up-selling
- Revenue enhancement through finding additional customers to up-sell

Customer Retention

- Current customer attrition/churn rate
- Revenue enhancement from predicting churn and targeting at-risk customers
- Revenue enhancement from real time, customer-facing actions to avoid defection
- Increasing customer lifetime value

Customer Acquisition

- Revenue enhancement through finding additional prospects
- Revenue enhancement through improving conversion rate of applications
- Improved efficiency in application rejection
- Improved efficiency in application processing
- Additional revenues from reuse of saved direct-marketing expenditures

Source: Gartner Research

Figure 5. Hard and Soft Benefits From CDI

5.3.1 CDI Benefits Case Study 1

Business Problem: Company A's profitability and market share were negatively affected by its acquisition of marginal, unprofitable customers. The enterprise could not identify its best customers because of its failure to weed out duplicate customer entries with name and address variations, thereby preventing it from monitoring its customers' buying patterns.

Solution: Company A implemented technology from a leading CDI vendor to integrate its customer data. Using the CDI product, the company was able to identify and weed out duplicate entries in its customer database, and create new revenue opportunities by identifying new customers.

Results: As these duplicate entries accounted for 12.5 percent of its total customer mailing list, Company A generated a total of $1.75 million in savings in mailing costs in a one-year period alone through the elimination of 5 million redundant mailings (see Figure 6).

Company A	Quantity/Amount
Total customer base	10 million
Annual mailings per customer	4
Annual mail volume	10 million x 4 = 40 million
Annual direct-mail costs	40 million x $0.35 = $14 million
Duplicate customers identified through CDI	12.5% of customer base
Duplicate mailings eliminated	12.5% x 40 million = 5 million
Savings in annual direct-mail costs	5 million x $0.35 = **$1.75 million**

Source: Gartner Research

Figure 6. Company A Cost Savings: Removal of Duplicate Mailings to Customers

In addition, the company was able to generate an estimated $33,000 in added revenue (less mailing costs) from mailings to new customers identified by the CDI tool (see Figure 7), for a total annual benefit in savings and revenue enhancements of $1.783 million. Forecast and compounded over a five-year period, this translates into potential hard benefits of more than $10 million, even before intangible (i.e., soft) benefits are included.

Company A	Quantity/Amount
New customers discovered through CDI	5,000
Annual mailings per customer	4
Total mailings to new customers	5,000 x 4 = 20,000
Cost of new mailings	20,000 x $0.35 = $7,000
Average customer response rate	2%
Average transaction per customer	$100
New revenue opportunity	20,000 x 2% x $100 = **$40,000**
New revenue less mailing costs	$40,000 - $7,000 = **$33,000**

Source: Gartner Research

Figure 7. Company A Revenue Enhancement: Additional Customers Recognized

5.3.2 CDI Benefits Case Study 2

Business Problem: Company B, a large telecommunications firm, embarked on a customer-centric approach to grow its business, but the incoming data from a variety of sources was anomalous and inconsistent. The company's multiplatform environment compounded this challenge: Billing was performed on legacy mainframe systems, with POS and data-warehousing systems running on Unix

systems. One of the company's goals was to standardize enterprise rules for data quality management, and share them across its established and future platforms. Company B needed a tool that was robust enough to process complex data anomalies and inconsistencies, as well as integrate online both with its existing POS and billing systems, and with the data transformation layer built for its data warehouse.

Solution: Company B selected a CDI product based on its ability to process input from multiple sources, and to re-engineer complex name and address issues. The product was applied to the billing system, which required ABC's team of business consultants and IT experts to analyze the billing system to formalize business rules.

Results: The billing system went live within six months, accompanied by a perceptible improvement in customer satisfaction. ABC was able to use the same logic set and business rules to provide a standardized, consistent view of its customers. The team also applied the knowledge it gathered to data in other customer applications.

5.3.3 CDI Benefits Case Study 3

Business Problem: Company C is a leading publisher with an annual direct-mail budget of more than $50 million. The company relies on customer data for direct-mail campaigns, mining customer segments, generating informational profiles and supporting campaign management systems. The business issues facing Company C include the high level of duplicate customer entries in its databases, and its inability to obtain a single view of its customers. Also, the company feels that the quality of its data, while good, could be improved.

Solution: Company C decided that a CDI product providing a single view of the customer would best enable it to meet its business needs. The company bundled the CDI product with other data enhancement and data-cleaning products to improve the quality of addresses. It was able to justify the investment of implementing the products based solely on cost savings.

Results: The company achieved savings of more than $1 million by eliminating duplicate addresses and duplicate names in databases, and suppressing mail sent to redundant addresses. This figure would be substantially larger if intangibles (e.g., customer satisfaction, increased revenue opportunity through recognizing additional customers and savings in operational costs) were considered. Company C expects to recover its investment within six to eight months of implementation.

6.0 The Care and Feeding of Customer Data

Clean data is a prerequisite for gaining the benefits of CDI. Thanks to the growth of data warehousing, and the fact that more companies view corporate data as an asset to be enhanced and leveraged, enterprises are less likely to take their data for granted.

The six major steps to cleaning data are:

- *Elementizing:* Parsing data into components, called "elements"
- *Standardizing:* Applying a standard form to these elements
- *Verifying:* Examining the elements and checking for errors
- *Matching:* Detecting identical elements, such as the same address or name
- *Householding:* Searching for information that indicates the same household

- *Documenting:* Capturing the results of the earlier steps in metadata to facilitate future data-cleaning exercises

6.1 Data Quality Tools

Data quality tools tend to fall into one or more of the following categories:

- Data analysis
- Name-and-address integrity
- Data re-engineering

Users can employ tools from more than one category to perform their data analysis and scrubbing. A variety of automated tools and technologies are available to assist in cleaning data during the extraction and transformation process. In addition, these tools can be used to reclean previously warehoused customer data.

Traditional data extraction and transformation tools are limited to providing data type and format (data syntax) transformation — i.e., they simply translate the data from one format into another with no interpretative adjustment. Newer, intelligent tools use business rules to transform the contents into new information. These tools are particularly important for integrating customer data, which, compared to product or financial data elements, tends to be more complex and to include more unstructured data (e.g., notation fields containing details of customer contacts).

- *Name-and-address-cleaning tools* essentially evaluate and clean name and address fields. Representative providers of these tools include Harte-Hanks and Acxiom.
- *Householding tools* identify relationships between accounts and customers, as well as customers' relationships with other customers and business entities. Representative providers of these tools include Acxiom and Vality Technology.
- *Generic data-cleaning tools* use various matching and rules-based routines to detect and correct data errors. Representative providers include Oracle, Innovative Systems and SAS Institute.

Some of these products — including all first-generation tools — are designed to do the cleaning on a batch basis, either before or after loading data to the target database. With the newest technology, most vendors support real-time data-cleaning processes. The technology used to ensure data quality typically employs "fuzzy logic" matching algorithms, address parsing algorithms and large look-up tables.

6.2 Data Quality and Third-Party Data Integration

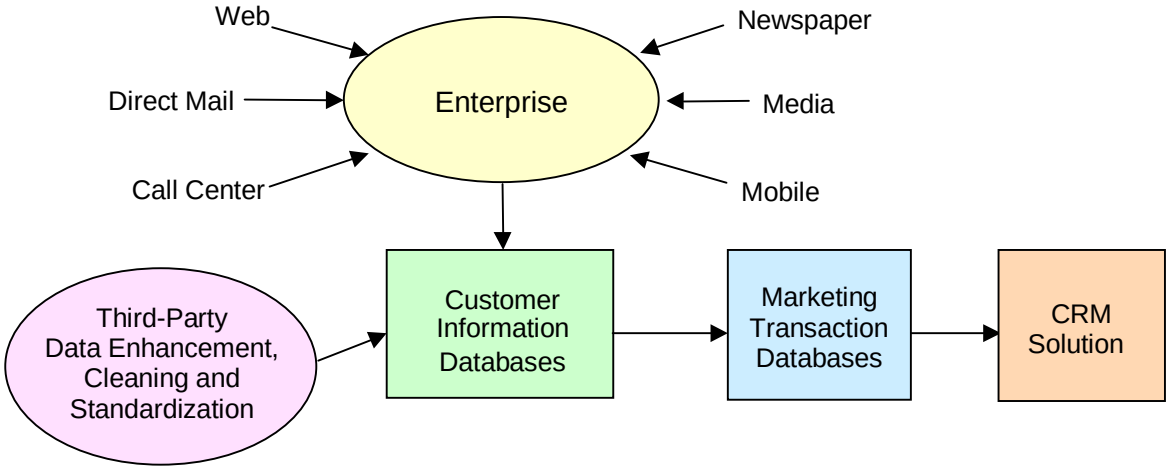
Internal data can be combined with data from third parties. Before an enterprise takes this step, it must ensure that:

- Internal data is cleansed and standardized into a format that can be appended.
- Business rules governing the target data are airtight.
- Transformation rules required for the data are in place.

When these conditions are met, the internal and third-party data can be combined to enhance the operational capability or usability of the data. Enterprises may turn to information suppliers for many different types of data, including:

- *Personal:* This includes information such as birth date and gender.
- *Geographic:* This includes data on latitude, longitude and time zones.
- *Postal:* National Change of Address (NCOA) and address standardization data are among the useful enhancements in this category.
- *Census:* This provides potentially useful data such as household income and makeup, and ethnic information.
- *Demographic:* This includes data on area reputation, population density and mean income
- *Economic:* Useful information in this category includes stock exchange data and historical price indexes.
- *Political and world event data:* Although such data may seem obscure, it could impact enterprises' business strategies, particularly for those trying to expand internationally.
- *Behavioral data:* Data such as marital status and religion is extremely useful for marketing professionals undertaking such arduous tasks as segmentation and psychographics.

As Figure 8 illustrates, every transaction is a possible source of data for a customer information database (CID). CID data must be cleaned and refreshed frequently because the quality of data entering these systems typically is poor. In general, this data is compared against a comprehensive database from a third party such as Acxiom, Experian or infoUSA.



Source: Gartner Research

Figure 8. Transaction Types Feeding Customer Information Databases

The quality of data in a CDI vendor's database drives the quality of information in the marketing transaction database, so enterprises should carefully consider vendor data quality when making a purchasing decision.

Because third party-data is often used to enhance other data, it is critical that enterprises understand:

- The meaning of the third-party data, and whether its context has changed after being appended from external sources
- The source of third-party data and its relative freshness
- The reliability of this data, and the enterprise's confidence level as to its accuracy

7.0 The CDI Technology Market

CDI is quickly becoming a buzzword in CRM circles, and enterprises and vendors are starting to talk about CDI as an established market segment. Many enterprises, however, are unclear about exactly what it includes.

Many see it as data-cleaning revisited, but a closer look shows it to be a key component that all CRM strategies should take into account. CDI, if implemented correctly, will reduce operational and marketing costs and enhance revenue-generating opportunities through increased customer satisfaction and the identification of new customers. In this section, we describe CDI's basic structure to define exactly what it comprises and the benefits it offers.

7.1 CDI Technology Defined

In some ways, CDI is nothing new, especially in the area of CRM known as database marketing. What is new is the integration of multiple capabilities into vendor products, providing ROI benefits by improving the underlying quality of the database. In addition, the emphasis that the creation of the CDI market has placed on customer data quality is a useful step forward in the CRM arena.

Gartner sees three main categories of CDI technology and services:

- Core technology
- Integrated components
- Services

The core technology elements represent the fundamental functional requirements of a complete CDI product. They are:

- **Hygiene:** This subcomponent improves data quality through address standardization and correction. Clean, accurate address data facilitates more-accurate integration and is a critical component of effective customer communication. The benefits it provides include:
  - Improving CDI success rates
  - Improving the validity of data downstream, which increases organizational confidence as CRM software and processes are rolled out
- **Linking:** This involves the matching of customer records as business needs dictate, including at the point of customer contact. Linking is the most prominent component of CDI, and full-function CDI products should be able to accommodate blended records (business and consumer), as well as multinational files in online and batch environments. Its benefits include:

  - Reducing duplication of undesired customer records
  - Identifying similar customer records within a database, across databases or across the entire enterprise

- **Grouping:** This refers to the ability to define different views of a customer based on specific business rules, and to vary these views based on business application needs. This may include an individual view, a household view, an address view or an account view. Grouping usually allows resident customer data — such as an account number, phone number or, as law permits, a social security number — to be used in conjunction with CDI linking routines to define unique business or situational views. Its benefits include:
  - Providing enterprises with proprietary control over the business rules used to define the customer views
  - Enabling separate views to be established based on department or application
  - Allowing data elements other than names and addresses to be considered when needed in the integration process
- **Customer Recognition:** This is the ability to accurately recognize customers based on identifying information, and to synchronize all internal customer keys and pointers on a continuous basis. Customer recognition includes data models, real-time recognition components, high-volume batch processing components and associated interfaces. Its benefits include:
  - Providing a pre-built infrastructure for continuous synchronization of internal customer keys and pointers
  - Enabling instant customer recognition and consolidation of disparate customer information

7.2 CDI Products and Vendors

Figure 9 lists a sampling of the products and vendors that offer core CDI technology as defined in Section 7.1. Note that the selection below is not intended to be an all-inclusive list, and any particular company's omission should not be interpreted as a negative reflection of its CDI capabilities or products.

Core CDI Technology Products	Vendors
AbiliTec	Acxiom
Trillium	Harte-Hanks
TruVue	Experian
Integrity	Vality Technology
Code-1	Group 1 Software
i.d.Centric	Firstlogic
DataFlux	SAS Institute
infoConnect	infoUSA/Donnelley Marketing

Source: Gartner Research

Figure 9. Selected CDI Products and Vendors

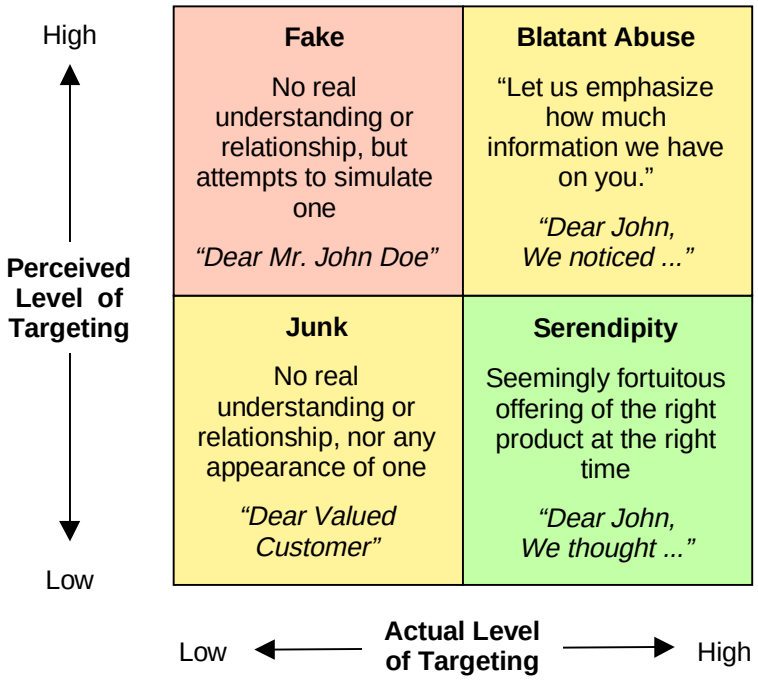
8.0 Customer Data Privacy Considerations

The purpose of CRM is to generate loyalty through customer-centric context, which data analysis helps bring to customer interactions. The idea is for corporations to build an affinity with their customers — however, in the process, they must walk a fine line between effectively targeting customers with custom-tailored offers and purchase suggestions, and "knowing too much" about their behavioral patterns. If customers fear that their privacy has been breached, the intended benefit of the CRM effort will be eliminated.

8.1 Privacy and Customer Interactions

An enterprise's approach to customer interaction is a key to successfully managing customer privacy. The optimal position in this regard is the "Serendipity" quadrant of Gartner's Privacy Matrix (see Figure 10). Under the doctrine of serendipity, the enterprise knows more than meets the customer's eye. This enables the enterprise to offer "just the right solution at just the right time" for a given customer, and to leave this customer with the impression that this fortuitous circumstance is the result of coincidence, luck or happenstance. Alternative strategies can alienate customers, either through poor targeting or blatant abuse of their privacy.

The Privacy Matrix has little to do with the amount of information collected, but everything to do with the customer's perception of how much has been gathered. The same amount of information could be collected in each situation; however the customer's *perception* of the personalization and information-gathering that has taken place is what varies. An approach falling in the "Serendipity" quadrant, for example, conveys the impression that the enterprise is simply offering a well-timed service. One falling in the "Blatant Abuse" quadrant, however, leaves the customer with the uncomfortable feeling that the company knows too much personal information — rendering future purchases unlikely.



Source: Gartner Research

Figure 10. Privacy Matrix

8.2 Privacy Hype

The hype surrounding online privacy has rippled beyond the Internet, and the heightened public concern about privacy manifested itself during the 2000 U.S. Census. When the U.S. Census Bureau began its count in the spring of 2000, it received more than 600,000 complaints concerning perceived violations of personal privacy. Many American citizens saw the census questions as too personal, even though many of these questions had been on the form for decades. Some citizens indicated they would rather risk fines than supply the information, thereby underscoring the perceived seriousness of their fears. The U.S. Congress had to step in to mediate the matter.

Corporate America could ignore this dispute; however, privacy backlashes have been occurring worldwide, and nothing is more likely to derail CRM — and negate the large investment therein — than concerns over privacy. Although from risk management and marketing perspectives, it would be easy to re-purpose off-the-shelf, security-driven solutions, Gartner would not advise this until the social and legal parameters of acceptable business behavior are more firmly established, and the degree of consumer acceptance is better understood.

Because the degree of privacy regulation varies in different parts of the world — from relatively lax in the United States to stringent in parts of Asia — enterprises must know the respective legislation within each market where they plan to implement a CRM solution. If an enterprise rolls out a worldwide CRM solution, it must monitor the customer information from each respective market differently.

Although the United States leads in CRM and Internet penetration, it is the laggard when it comes to customer privacy and government regulation. Gartner believes that the United States will eventually move toward the European standard, where the customer owns their information, not the corporation. Organizations embarking on CRM solutions need to be aware of this issue and plan accordingly.

8.2.1 Action Items

Gartner offers the following recommendations to enterprises with regard to privacy issues surrounding their customer data:

- Be aware of the sensitivity of current and potential customers.
- Be aware of the varying degrees of regulation around the world, and be prepared to adhere to more stringent standards in foreign markets.
- Establish a privacy policy.
- Develop a crisis communication program.
- Establish privacy as a value proposition, and as a corporate priority.

9.0 Conclusion: The Road to High-Quality, Integrated Customer Data

A clean, well-integrated customer data repository is a prerequisite for CRM success, as it is for any data mart, data warehouse or a data store. Integrating customer data is a challenging task, however, because merely assembling the data in one place is not enough. A number of steps must be performed, including data extraction, transformation, cleaning and consolidation.

9.1 The Challenge of Data Collection and Integration

An enterprise may choose to combine its CRM efforts with a data-warehousing project, or it may instead choose to establish a customer data mart specifically targeted to the CRM application. Either way, this requires integrating customer data from disparate databases within the enterprise. These databases are likely to be located in various parts of the enterprise, collected from varying sources and channels, and stored in a multitude of differing architectures and platforms.

In attempting to deliver a centralized customer information repository, many enterprises underestimate the challenges of data extraction, transformation and cleaning. At least 50 percent of enterprises undertaking a CRM strategy are unaware of data quality problems in their environment. The remainder recognize the existence of data quality problems but do little about it due to a lack of perceived value.

The resources required to perform the cleaning and consolidation of customer data depend on the size and complexity of the environment (i.e., the number and types of data sources, data volumes and the prevalence of data quality issues). The physical extraction and integration of this data may be achieved through extraction/transformation technology (see Section 6.1). At least 80 percent of enterprises underestimate the time and resources required, with many exceeding their budgets for this effort by two or three times.

Enterprises that fail to address data quality issues risk missed opportunities and operational inefficiencies. Although it is rarely a trivial undertaking, developing and maintaining a high-quality, integrated customer data repository is worth the effort. It is the means to achieve the important benefits described in Section 5.2, namely:

- Cost savings from the removal of redundant customer data
- Increased revenue from identifying and targeting first-time customers
- Enhanced revenue from higher customer satisfaction and retention
- Savings in operational costs

9.2 A Partnership of Business and IT Resources

Avoiding data quality pitfalls and achieving these benefits requires analysis of all customer data sources, and the definition of mapping and integration approaches to consolidate this data into the customer data repository. Significant involvement of both business and IT resources is required to identify and define corrective action for data inaccuracies and discrepancies.

To assist in this effort, enterprises may opt to deploy one or more of the various styles of data quality technology, but decisions on how specific issues should be resolved must be driven by business leaders (see *Research Note* COM-13-5615, "Data Warehouse Acceptance: Cultural vs. Technical Issues").

To ensure that their customer data is of high quality, IT and business resources must:

- Collaborate in setting standards for operational and CRM applications
- Deliver data-related training and procedural improvements
- Foster a culture where data is treated as a valuable asset

High-quality customer data is necessary to achieve CRM benefits. Without adequate resources to focus on data quality, an enterprise's CRM efforts will not be successful. IT professionals must work with business users to identify and raise the visibility of data quality issues

9.3 Summary and Recommendations

Gartner's high-level conclusions regarding data quality, CRM and CDI include the following:

- To appreciate the effectiveness of a CRM solution, enterprises must understand how three information types — descriptive, relationship and contextual data — interact to provide a coherent picture of the customer relationship.
- Building an understanding of the customer relationship requires different types of analysis of customer data, depending on how an enterprise will eventually use this data. A successful CRM strategy dictates that both analytical and operation CRM be integrated, and that organizations understand the difference between segmenting customers (analytical CRM) and targeting them (operational CRM).
- An effective CDI solution is a vital component of CRM. CDI can best be described as the combination of technology, software, processes and services required to achieve a single, accurate and complete view of the customer across multiple sources of customer data (internal and external), databases and business lines.
- Developing and maintaining a high-quality, integrated customer data repository is a prerequisite to achieve the cost reduction and revenue enhancement benefits described in this *Strategic Analysis Report.*

In light of these conclusions, Gartner offers the following recommendations to enterprises seeking to gain the benefits of high-quality, integrated customer data:

- Incorporate real-time data into decision-making models. These models will in turn trigger optimized, automated responses, or provide the end user with suggested courses of action during normal operations.
- Stop thinking in terms of individual channels (e.g., direct mail or the Web) when communicating with customers. To effectively serve their needs, marketing organizations must present customers with a

unified set of offerings and a quality service experience across channels — enabling customers to deal with the enterprise when they want, how they want and where they want.

- Because third-party data is often used to enhance the enterprise's customer data, it is critical that organizations understand:
  - The meaning of the third-party data and whether its context has changed after being appended from external sources
  - The source of the data and its relative freshness
  - The reliability of the data, and the enterprise's confidence level as to its accuracy
- Investigate the capability of CDI products to provide ROI benefits by improving the underlying quality of the customer database. If implemented correctly, CDI can reduce operational and marketing costs and enhance revenue-generating opportunities through increased customer satisfaction and the identification of new customers.
- With regard to data privacy issues, enterprises should:

– Be sensitive to the privacy concerns of current and potential customers

  - Be aware of the varying degrees of regulation around the world, and be prepared to adhere to more stringent standards in foreign markets
  - Establish consumer privacy as a corporate priority, establish a privacy policy and communicate privacy as a value proposition to customers.

- To ensure that their customer data is of high quality, IT and business resources must:
  - Collaborate in setting standards for operational and CRM applications
  - Deliver data-related training and procedural improvements
  - Foster a culture where data is treated as a valuable asset

Appendix A: Acronym Key

AI	Artificial intelligence
CDI	Customer data integration
CID	Customer information database
CMS	Campaign management system
CRM	Customer relationship management
DW	Data warehouse
ERP	Enterprise resource planning
GUI	Graphical user interface
ISV	Independent software vendor
IT	Information technology
NCOA	National change of address
ODS	Operational data store
POS	Point of sale
RFM	Recency, frequency, monetary
ROI	Return on investment
SI	Systems integration