



# Data Quality And Integration In Banking

Most companies have an enormous amount of customer data . A typical bank has over 500 million data elements per \$1 billion in assets. Even with the best software applications, the quality of customer data decays over time. This is not due to the applications, but to the nature of the data itself. Consumer and business data changes frequently and sometimes dramatically. The quick rate of change in key identifying elements of customer data makes the problem worse. People move, change names, and change phone numbers.

Data often sits in isolated systems that are not linked together. In retail banking the Customer Information File or Customer Information System has been used for years to attempt to link important customer data. However, these systems rely on end users to keep the data clean and linked. The US Postal Service estimates that 40% of the data keyed by users is either incorrect or incomplete. Many banks have core processing systems (mortgage, credit card, and brokerage) that are not even linked into the CIF/CIS.

This lack of data quality and the complicated nature of customer to customer, customer to account, customer to household, and customer to business relationships leads to a major roadblock in achieving business objectives: **an incomplete view of the customer.**

## DATA QUALITY – A BUSINESS PROBLEM

---

Most IT and business people understand that poor data quality is a business problem. Studies show that up to 25% of data in an average bank's CIF/CIS is incorrect. Errors in customer data lead to numerous issues that impact a bank's bottom line. Some of these problems are easy to understand and measure. For example, incorrect addresses increase mail costs for a bank. Even for a smaller bank that sends out 200,000 promotional mail pieces a year, a 25% error rate would account for \$50,000 in mail costs.

The expense and lost revenue resulting from poor data quality is often more subtle. Data quality problems start a chain reaction in a bank's business processes. For example, bad data makes privacy violations much more likely. Contacting the wrong person the wrong way can cost the bank up to \$10,000 per incident. On a larger scale, data quality problems lead bankers to make the wrong decisions both tactically and strategically, resulting in lost customers and decreased profitability.

## WHAT'S DRIVING DATA QUALITY?

---

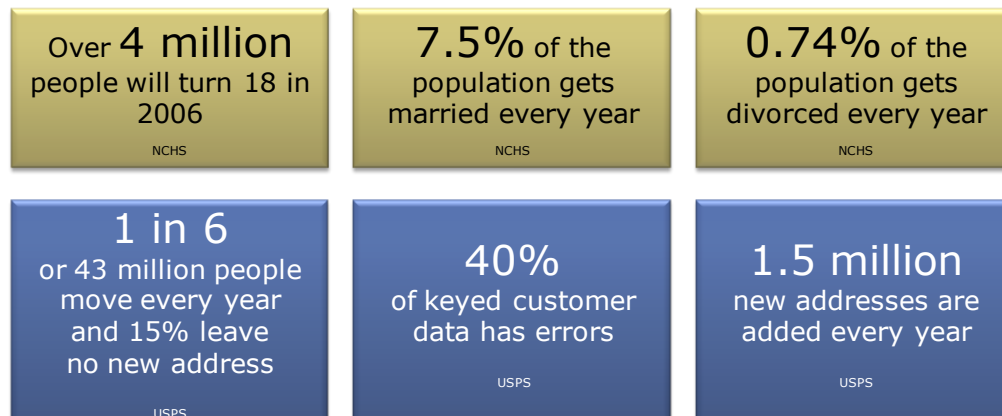
In many ways the high profile of the data quality issue is due to the banking industry's need to provide differentiating customer service, better marketing practices, and better customer value. The large number of failures of CRM projects has pointed out the obvious need to fix data quality problems. Banks of all sizes have struggled with the implementation of front-office CRM tools, marketing CIFs, business intelligence tools, and analytical systems all because of data quality issues.

Data provides the cornerstone to build these capabilities and achieve business goals; therefore, the highest quality of data possible is required. Part of this cornerstone is the ability to see a complete view of the customer. To understand the full value of a relationship with a customer, you must be able to link all accounts that belong to an individual as well as identify other relationships that might exist in the household or across business accounts. Additionally, you must know who within the business structure owns the customer, where that customer does most of his or her business, and what products that customer has. Unfortunately, name and address information, which provides the key to linking across accounts, is often entered inconsistently during account set up, and detailed information about who owns the account or what products the account contains gets buried in multiple, unstructured fields along with other critical pieces of information.

## MOVING TARGET

---

So, if data quality is so important, why haven't banks fixed the problem? While it may seem like a simple problem on the surface, data quality has been more of an art than a science. The nature of the primary problem, names and addresses, makes fixing the problem much more complex than it appears. Name and address data is quite unstructured in most systems. The address lines alone can contain a huge variety of information about a customer. This data is almost always manually keyed into bank systems leaving plenty of room for typos, transposed characters, and entry of data into the wrong fields.



It's also important to understand that the data is a moving target. Customers constantly change key information when they move, marry or divorce, and change names, phone numbers, or email addresses. Complexities in address information like street aliases and changes in ZIP codes complicate the problem even further.

## THE 360<sup>0</sup> VIEW

---

For years people in the financial services industry have talked about the need for a "360<sup>0</sup> view" or "single view" of the customer. Whatever it is called, the objective is to present a complete, consistent, and correct picture of customers and businesses to all areas of an enterprise.

While this is a very simple concept, it is extremely difficult to implement due to data quality issues and a lack of understanding of the requirements. A single view of the customer can only be achieved with clean data and a mechanism to link different customers and businesses together.

Bank Customer Information Files or Customer Information Systems represent the oldest strategy for achieving this linkage. This strategy largely depends on individual users making those links as they sell and service accounts. Over time, it has become clear that this approach has resulted in error rates of

more than 25%. These customer systems rarely have all of a bank's relationships represented. While the customer system has core deposit and loan relationships defined, it rarely has data on mortgages, credit cards, trust accounts, investments, or insurance products.

### Complete

The 360<sup>0</sup> view of the customer must be complete. That is, it must have all relevant data about the customer. Many banks calculate the profitability and value of a customer relationship based only on core products in the customer information system. But what if the customer has a trust or investment relationship with the bank? Shouldn't all of the relationships be considered?

### Consistent

The 360<sup>0</sup> view must also be consistent. Everyone must be looking at the customer in the same way. From analytical processes to retail delivery channels, the same information about the customer must be used. What if all of the relationships were considered for profitability analysis, but the customer service representative in the branch could only see a fraction of those relationships?

### Correct

Finally, the 360<sup>0</sup> view of the customer must be correct for the specific business process or person. While we usually speak of a "customer view", a particular business process may need to look at a wider perspective. For example, in marketing processes it may make more sense to use a 360<sup>0</sup> view of a household. In fact, there may be requirements for several different views each of which must be complete and consistent.

### Constructing the 360<sup>0</sup> View

The key to constructing the 360<sup>0</sup> view lies not in a particular application but in the data itself. Customer identity data is the lowest common denominator across all data sources. Name of the consumer or business, physical address, telephone numbers, email addresses, SSN or Tax ID, birth date, and driver's license numbers can all be used to create the 360<sup>0</sup> view. This is far from a trivial process. Each of these data elements has its problems and they are all notorious for having dirty data.

The most sophisticated and accurate techniques for using identity data involve using all of the elements in concert. The name and address fields from banking applications also contain important details required to link individual customers and businesses. Locating and using relationship indicators such as "and", "or", "doing business as", or "custodial" are vital part of constructing the 360<sup>0</sup> view.

## Using the 360<sup>0</sup> View

It is very important to realize that the 360<sup>0</sup> view is only a means to an end. Achieving this view is a vital step to using customer data wisely, but it is only one step. The 360<sup>0</sup> view must be used in downstream processes to truly affect business performance. Sales, customer service, profitability analysis, relationship banking, fraud, and risk process can all be dramatically improved through this complete, consistent, and correct picture of the customer.

## WHAT'S REQUIRED TO FIX THE PROBLEM?

---

Fixing data quality and integration issues requires sophisticated software components, business rules, and serious expertise. The following description of the process details the requirements for achieving the high data quality.

### Profiling

The first step in addressing data quality problems is to understand the data content. The Profiling process looks at each data source and provides a profile of that data. The data profile shows how many records and elements are in the data source and information about each field. This includes the unique values present in each field, maximum values, minimum values, and other attributes that describe the data.

### Validation

Validation sets a baseline for standard information for each data element. It determines if a data element meets those standards and sets thresholds and rules for what to do if data does not meet the expected standard. In some cases validation uses a list of acceptable values for each data element to determine the validity of the incoming information. For other data elements, especially numeric fields, ranges of acceptable values are used to check the data integrity.

When data does not meet the validity standard, several different actions may be taken depending on the applicable business rules. The data may be mapped to a default, grouped as an unknown value, or even rejected if the discrepancy warrants it.

### Parsing

Parsing examines unstructured data and decomposes it into specific components. This is especially useful for name and address data. Names must be parsed into prefix, first name, middle name, last name, and suffix components. The different ethnicities of names make this particularly difficult. In a similar fashion, address data must be broken into its components.

### Standardization

Standardization looks at individual data elements and puts them into a standard representation. For example, the suffix “Jr” gets set as the accurate and standard abbreviation for “Junior.” Address elements are standardized according to postal regulations. Another example includes simplifying address information across core processing systems to standardize spelling and abbreviation for cities. For example, a single data source can have three variations of the same name entered by three different users: “Saint Paul”, “St. Paul”, and “Sant Paul”.

The same lack of consistency applies to other information in core systems such as dates, telephone numbers, driver’s license number, SSN/TIN, and account numbers. To perform accurate data integration, all of this information must be represented consistently.

### Transformation

Data must be transformed at both the element and the structure level to be more suitable for downstream processes. The transformation process can look across data elements to fix additional errors. For example, an intelligent comparison of SSN with date of birth and geographical data can spot consistency problems. The transformation function may also split or concatenate data.

### Enhancement

In many cases, the standardized and transformed data can be correct, but still missing key information. Enhancement processes can add information that is missing from the original records. For example, salutation, gender, and other elements can be added based on name data. Addresses can also be enhanced by adding street direction and even apartment number.

### Audit

Cleansing processes are vital in meeting the overall business requirements and ensuring high quality data. To validate that the cleansing process is working correctly, all of its steps must be auditable. Auditing allows business users as well as processors to understand where the data comes from, where it goes, and how each function massages the data to make it higher quality. This is also known as “data lineage.”

### Integration

After the previous steps are complete, data must be constantly updated and integrated into the views necessary to support other business processes. This ongoing integration requires an understanding of relationships often hidden in the data itself.

Integration builds the foundation of the complete customer view by linking previously unconnected information such as accounts held by the same individual or accounts held by different individuals

within the same household and tying this information to additional prospect, demographic, and lead information from external sources.

## Matching and Grouping

Traditional data integration processes often bring all data together at a single point in time and then sort that data into groups of like records. These records are then merged together to form integrated records for specific purposes. When the integrated records must be updated, either with new data or updates to existing identity data, all of the data must be brought together again in another massive, time consuming effort. This process often leads to data that is stale before it is used, almost impossible to update in a timely fashion, and often difficult to use for other purposes. It also ignores the intelligence gained from integration in previous periods that can significantly enhance the accuracy of integration.

## Recognition

To solve the data quality problem and accomplish the bank's larger business goals, the solution must go beyond matching and grouping to be much more flexible and dynamic when it comes to integrating data. Because data will arrive at different intervals and in varying quantities—from single records arriving continuously out of customer interactions in real time to larger batch files sent by core processing systems—a different style of integration is required.

Recognition works on a principal of learning about customer identities and relationships as data is processed. From the initial load to every additional transaction, the system learns and extracts meaning from the customer elements in the data to constantly enhance integration capabilities. This dynamic, learning nature allows recognition functions to locate more complicated but often more valuable relationships across multiple lines of business in an organization. Many integration systems use the primary information listed on accounts and do not look at secondary relationships to create links with other accounts across the business. A recognition component improves this model, for example, by using a custodial relationship on a trust account and linking that to checking and savings accounts.

The Recognize function takes identifying information from a record and uses that information to determine if there is a matching individual or household already on a master index. Identifying information includes elements such as name, address, phone number, account numbers, date of birth, driver's license number, SSN/TIN, and external identifying data. This function must use "fuzzy" matching techniques to identify close matches in character data.

Recognition also supports real-time access by channel applications to locate customer information if it is not found using traditional core system functions. The system supplies a persistent ID for each individual found; or, if the individual is new, adds that individual to the cross-reference and creates a new ID.

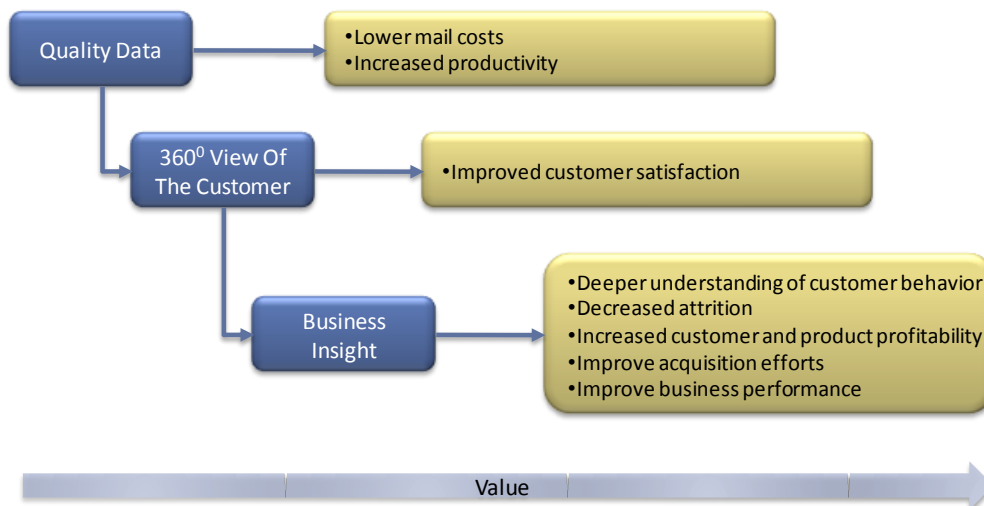
Core systems often use exact field matching to find individuals, which means that users in the front office can often miss locating an individual's information when only that individual information is available. The index that is created by recognition processes cross references individuals on a variety of information such as name and address, account number, SSN, and phone number so that no matter what information they enter, front office users will get back a complete view of the customer.

### Data Stewardship

Banks must understand the various ways customer data is stored and create data standards driven by business rules to catalogue and map information that spans multiple core systems, channel applications, and external data sources. This process, known as "data stewardship", treats data as a bank asset to ensure that data is used consistently and accurately.

## THE VALUE OF QUALITY, INTEGRATED DATA

The process outlined above can improve the accuracy and value of data. While the process can be used to clean up a bank's CIF/CIS, the value of clean, integrated data goes well beyond this. For example, clean data can be used to reduce expenses in the bank by reducing postage and the amount of returned mail and by increasing productivity of back-office staff. However, the value of quality data extends far beyond these initial savings.



Using quality data to produce a 360<sup>0</sup> view of the customer in day-to-day interactions with customers can lead to much higher customer satisfaction. Using that same integrated view, downstream business processes that give insight into customer behavior can deliver value many times that of just saving mail costs. For example, contacting a customer or prospect in the wrong way during marketing campaigns could cost a bank \$10,000 per incident. Using this insight to decrease attrition can save millions of dollars in lost contribution.

## WHAT ARE THE ALTERNATIVES?

The data quality problem for banks is not new. For years, banks have engaged third-parties to perform “scrubs” on their CIF/CIS data file. However, most experts agree that this approach to the problem only ensures name and address quality at a single point in time. More modern solutions focus on frequent cleansing of not just name/address data, but of multiple data elements that can support downstream business processes. Top US banks are moving toward a continuous process that provides daily and even real-time cleansing and integration. The following table summarizes some of the alternative available to banks today:

Alternatives	Strengths	Weaknesses
<b>Service Bureaus</b>	Experience and history in cleansing name and address data for direct mail.	Offsite processing of data introduces long wait times in what should be a continuous process. They only provide part of the solution.
<b>Data Integration Tools</b>	Sophisticated software tools for data quality.	The implementation and integration of the tool can be very expensive. Onsite data integration experts are needed to run the system on an ongoing basis.
<b>Customer Hubs</b>	These tools provide a modern CIF/CIS approach to house one “gold” copy of customer data.	The root data quality issue remains. Modification of existing systems must be made to retrieve customer data from the hub.
<b>Professional Services</b>	Highly customized approach for large banks.	These solutions can be extremely expensive with a minimum cost of \$5 million per year.

## CONCLUSION

---

Data quality and data integration are two important topics for today's financial services companies. The lack of quality, integrated data is a serious business problem that can have a substantial impact on a bank's financial performance. That is why large banks have spent millions of dollars on data quality and integration efforts. These projects are supported by bank executives because they understand the effect that good data has on key business processes. Smaller banks should partner with seasoned specialists who understand the problem to see similar benefits.

## ABOUT INSIGHT ECOSYSTEMS

---

**Insight Ecosystems** is an independently-owned, Arkansas-based customer relationship management and business intelligence company. The company was founded and is staffed by industry innovators who served as senior executives at Fidelity Information Services and at Acxiom Corporation. From our collective experiences and extensive knowledge of banking information technology, we have created a unique system that solves business challenges that bankers have faced for decades.

**Insight Ecosystems** provides clients with a business intelligence ecosystem that learns, grows, and changes to meet their ever-expanding needs. **Insight Ecosystems'** services and solutions empower companies to gain tangible insight into their customers, products, and financials, turning **data into insight** and **insight into action**.



**Insight Ecosystems**  
16101 LaGrande Dr.  
Suite 100  
Little Rock, AR 72223

501-448-0240  
501-448-0166 (Fax)  
<http://www.insightecosystems.com>  
[info@insightecosystems.com](mailto:info@insightecosystems.com)

### in-sight e-co-sys-tem

n. An environment of complex data processed within an analytical system that derives underlying, interdependent relationships and clearly communicates business knowledge that is immediately useful and extremely valuable.