



The Unfortunate  
Little Secret About Current  
CRM Data Cleansing.

(And how it destroys  
your bottom line.)

■ *Until now clean data was more myth than fact.*

That's because there is a crucial difference between ordinary CRM data de-duplication software and the next generation of data cleansing and standardization software offering [Contextual Data Shaping™](#).

*These next generation [Contextual Data Shaping™](#) Solutions do more than just clean your data.*

They take data cleansing one important step further than the current standard by automatically processing data in context to correct additional mistakes and eliminate duplications. So you get data that's instantly useable and ready to monetize. All without wasting your staff's time or your money.

*Even if you're cleaning your data as it enters your company...*

Your own sales, marketing and customer support people, to name a few, are changing your customer data. What was good customer or prospect information yesterday, has now been run over by new data that's been imported by functional groups, from the web and outside sources. All these interactions are opportunities to corrupt.



■ *That's why next generation **Contextual Data Shaping™** Solutions have been specifically designed to help you address the following areas:*

- ROI protection and improvement
- Smarter decision making based on accurate customer information
- Error correction leading to fewer mistakes
- Relationship building with clients
- Reputation Protection: Both yours and your company's

■ *Making smarter decisions*

A comprehensive data quality program is the most essential initiative to accurate customer information that allows you to make smarter decisions. Many events mandate improving data quality. These events, if not effectively managed, drastically deteriorate the value of your CRM investment.

“The high statistics of CRM failures have very little to do with the CRM itself. They have to do with companies fundamentally misunderstanding their information: what data they have and where it comes from”.

–Jill Dyché  
*The CRM Handbook*  
(Interview excerpt, CIO Magazine)

## *Some of these events include:*

- Deployment of a new CRM system or any sales or marketing automation system
- The addition, removal or update of information by employees
- Area codes and zip code changes
- Non-standardized formats
- Executive mandate for more accurate reporting
- A merger or acquisition
- New leads generated from a trade show or other event
- Import of demographic and psychographic information

## *Evaluation and Understanding*

It is vital that an organization understand its data – its sources as well as its manifestations. It is imperative to be able to view data in different orientations, in order to determine which source is most reliable. One common example is determining which elements of records must be retained in order to create the “ideal” customer representation. The bottom line is that when the requirement for consolidating customer information arises, as it inevitably will, rules must already be in place to define relationships, content, quality, and structure.

## *ROI protection and improvement through **Contextual Data Shaping™***

CRM installations are very costly and the data used with the CRM installation must be actionable in order to fully monetize your investment. That's why the ideal technology must be application- specific and versatile enough to work with both on-premise and hosted CRM applications. Since every CRM application has its own context, the ideal solution would connect to a variety of CRM applications and automatically recognize the requirements of each system, thereby making your data relevant, actionable and ready to monetize.

### Protecting your data over time takes maintenance.

The best laid plans for creating the highest quality of data will be for naught if standards are not implemented for on-going quality purposes. Although there are no truly fail-safe procedures, particularly when data is constantly being consolidated across applications and other sources, automated filters can reduce the degree of corruption that enters the CRM system. Periodic complete refreshing of all data in the system should be a high priority for every organization.

## Superior Data Verification through *Contextual Data Shaping™*

Up until today the industry standard for Data Quality has focused solely on Data Cleansing.

An ideal **Contextual Data Shaping™** Solution would allow both your technical staff and your non-technical staff to cleanse data and put it through a Contextual Screen™ every time they touch the data. That way data can be put into any shape that the user desires with confidence that it is not only clean but also usable.

Keeping data clean requires correction of errors, such as spelling, capitalization, or even incorrect information like zip codes and area codes. Mistakes in customer information can waste the time of mission-critical staff.

Data Verification Example: City Spelling Correction

	Account Name	Contact Full Name	City	State	Zip Code	Country
Before	AbbottLaboratories	Lou Balbo	Harisbrg	Pa	17101	USA
After	AbbottLaboratories	Lou Balbo	Harrisburg	PA	17101	USA

Data Verification Example: Zip Code Verification

	Account Name	Contact Full Name	City	State	Zip Code	Country
Before	Arad PerformanceCo	Amy Turner	Concord	Ma		USA
After	Arad PerformanceCo	Amy Turner	Concord	MA	01742	USA

**Contextual Data Shaping™** corrects mistaken information because

it understands the context regardless of the application, something you can't rely on with current data cleansing solutions. And unlike single point solutions, [Contextual Data Shaping™](#) works across a full spectrum of functionality.

## *Data Standardization*

Industry standard and organization-specific words and terms cannot be verified using common reference sources, such as a spell-checker. It is important that accurately maintained directories be available to cross-reference specialized terminology.

Data Standardization Example: Fortune 1000© Business Name Directory & Addressing Conventions

	Account Name	Contact Full Name	Address 1	Address 2	City	State	Zip Code	Country
Before	AMD	David Scardino	2001 King of Prussia Road, Suite C		Austin	TX	78728	USA
After	Advanced Micro Devices, Inc.	David Scardino	2001 King of Prussia Rd	Suite C	Austin	TX	78728	USA

	Account Name	Contact Full Name	Address 1	Address 2	City	State	Zip Code	Country
Before	Coca-cola	Robert Smith	Bldg. 900 Room 4	4206 Grand Ave.	Austin	TX	78728	USA
After	The Coca-Cola Company	Robert Smith	4206 Grand Ave	Bldg 900 Room 4	Austin	TX	78728	USA

## *Don't irritate your clients*

You work too hard to acquire and retain clients. That's why you want to avoid the kind of problems that can be caused by dirty data. You don't want to bombard a key contact with too many emails or calls or insult them by getting their information wrong. Instead, you want to look smart, organized and professional with every contact. That's why you need [Contextual Data Shaping™](#) and all it offers.

### Data Enhancement

Enriching the CRM user's experience often requires enhancing existing data. Types of enhancements include appending external information to existing data, or even filling in missing information. Common examples are missing zip codes, D&B data or industry-specific information (e.g. IMS). Failure to take advantage of this information reduces the value of a CRM system by preventing the organization from getting a complete picture of each customer.

### Duplicate Identification

To create the best view of a customer, there must be just one view of that customer. The best methods for duplicate identification use the knowledge of the organization's business process to define rules for merging disparate data records. The best tools have the flexibility to allow the user to change rules as needed, allow for easy configuration, and provide immediate feedback of results. Successful deployment of these tools can save significant time and money, as compared to the labor-intensive, error-prone method of manual cleansing.

Duplicate Identification Example: Match of Account Name, First Name Misspelled

Dup?	Account Name	Contact First Name	Contact Last Name	Address 1	Address 2	City	State	Zip Code	Country
1.1	Massachusetts Life Insurance	William	Elliott	300 Church St	Dept C	Austin	TX	78728	USA
1.2	Nassachusetts Life-Insurance	Wi1liam	Elliott	300 Church Street	Dept D	Austin	TX	78728	USA
Final	Massachusetts Life Insurance	William	Elliott	300 Church St	Dept C	Austin	TX	78728	USA

Duplicate Identification Example: Match of Account Name, First Name Long Form vs Nickname

Dup?	Account Name	Contact First Name	Contact Last Name	Address 1	Address 2	City	State	Zip Code	Country
2.1	Klaus Productions	Becky	Shelton	200 W Airpak Drive		Moline	IL	61265	USA
2.2	Claus Productions	Rebecca	Shelton	200 W Airpak Dr		Moline	IL	61265	USA
Final	Klaus Productions	Rebecca	Shelton	200 W Airpak Dr		Moline	IL	61265	USA

*Is it worth risking your company's reputation and client relationships on dirty data?*

We don't think so and neither do you.

So, while many companies have attempted to solve their data quality problems using various shortcuts it's definitely not something we recommend. Inevitably the captured value and quality is reduced, the understanding of data is insufficient, and risk is greatly increased. The following "misconceptions" often arise, only to be discarded after failure, with plenty of wasted time and effort.

## 1. Custom Scripts:

*"We'll just ask our IT staff to write scripts to clean up the data."*

There are multiple dangers associated with this direction since your data changes constantly, as it is used throughout your organization:

### Testing

The effort and time invested in debugging and testing scripts is usually more than the actual design and programming. And with mission critical data quality work it's impossible to keep up with all the changes in data.

### Scalability

Custom scripts typically are not engineered to scale either in sophistication, as your data quality needs grow, or in volume. Often scripts can work on smaller, initial data sets, but as the data sets grow in volume and complexity, scripts will fail.

## 2. Manual:

*"We'll just clean up the data ourselves by hand."*

Hand cleansing of data can be accurate for a small organization where only

one person is in charge of a small amount of customer data. Larger organizations that spend time manually cleansing often find that the ongoing costs to maintain data accuracy are prohibitive. Unfortunately, these hidden costs are often realized too late.

### — Standards

Each person will use slightly different standards to process the data and the ability to enforce one standard, manually, can be extremely difficult and in most cases practically impossible.

### — Repeated cleansing

Manual cleansing is only good for a limited period of time and must be done repeatedly.

### — Difficulty

Some items are very difficult to spot manually; for instance, identifying that “Mass” and “Nass” are duplicates. These entries do not show up close to one another in the database because of the spelling of the name.

### — Cost

It is too expensive to achieve enterprise-wide data quality using hand cleansing. Even cleansing thousands of records, on a continuing basis, is too expensive. Very quickly, staff fatigue and boredom will begin to show its mark – reduced quality and mistakes. Eventually, the company will end up having to pay to fix even more new errors without even the original task being complete.

## *It's Time for your data to come clean*

It's time to put the secret of clean, contextualized, useable data to work for you and your company. It's time to base your decisions on data you can trust and act on immediately. It's time to discover the benefits of next generation data cleansing with **Contextual Data Shaping™** Solutions.

### *About ActivePrime*

ActivePrime is the industry leader in **Contextual Data Shaping™** with five patents now pending. The ActivePrime CleanCRM CDS™ Solution is the only certified data quality solution for Oracle's Siebel CRM On Demand. ActivePrime is also certified for use with Salesforce.com and Sage SalesLogix.

### *For complete information contact*

[Info@ActivePrime.com](mailto:Info@ActivePrime.com)

### *About the Author*

Rosaline Gulati is an ActivePrime co-founder and President. She has worked with hundreds of companies on issues of deploying data quality solutions with CRM systems. In 2006, she was selected to chair the MIT CIO Symposium. Rosaline has an MBA from MIT Sloan, an MS Electrical Engineering (Technology Policy) from MIT and a BSEE from University of Illinois, Urbana-Champaign.