

# OUR DATA IS A MESS!

HOW TO CLEAN UP YOUR MARKETING DATABASE

BY BERNICE GROSSMAN & RUTH P. STEVENS  
OCTOBER 2005

# OUR DATA IS A MESS! HOW TO CLEAN UP YOUR MARKETING DATABASE

## EXECUTIVE SUMMARY

Business-to-business companies are often frustrated by inaccurate customer information. But there are steps you can take to keep your data clean and up to date. The most essential action steps are manual, using processes to enter data correctly in the first place, and to conduct outbound communications to verify its ongoing accuracy. These steps can then be supplemented by the automated method, which usually means sending your data to an outside service provider for regular clean-up. The authors sent a sample of 10,000 business records to four leading vendors and share the results.

---

### “WHAT DO YOU THINK ABOUT YOUR CUSTOMER DATA?”

You can ask this question of anyone involved in B-to-B sales and marketing, and the answer you receive will just about always be the same. “Our data is a mess.”

Of course, it’s likely that the answer will be couched in more forceful terms than “a mess.” But the implication is clear. People in business-to-business marketing are aware that they need to do a better job of collecting and maintaining accurate and up-to-date information about their customers and prospects.

There are steps you can take to keep your data clean and fresh. This paper reviews the data hygiene methods available to business marketers today. It will also introduce the results of a research study among data hygiene vendors that will help you understand what you can expect a third-party service provider to do to keep your customer information clean.

### WHAT IS A “MESS”? THE STATE OF YOUR BUSINESS- TO-BUSINESS DATA TODAY

Part of the problem faced by business marketers is definitional. While everyone says “My data is a mess,” they may mean different things by it. Marketers, for example, may be talking about situations when their direct mail arrives but doesn’t get delivered beyond the mail room. For sales people, it’s when they pick up the phone and discover the customer’s direct phone number has changed. But there’s more. The business may have moved its offices. Or the customer’s title may have changed. Or the data fields may be mixed up, for example, an old purchase order number that’s parked in the customer name field. Or the state of Nebraska may be abbreviated NB, while the post office only accepts NE. It goes on and on.

Each of these problems is common in business marketing databases, and creates enormous waste — of marketing communications investment, and of business opportunity — not to mention frustration

at all levels. So what can you do about it?

The solutions lies in data hygiene, defined as follows: Correcting inaccurate fields and standardizing formats and data elements. There are two general approaches to data hygiene: manual options and automated clean up. Let us look at what each of these can — and cannot — do for you.

## MANUAL PROCESSES

There are two key manual methods involved in data hygiene:

1. Enter clean data in the first place
2. Institute on-going updating processes

The most important is the first: If the data is entered or received incorrectly at the start, you have not only wasted a business opportunity, you have created needless extra expense to go back and correct the information. Bad data is worse than no data at all.

Smart companies are using the following key methods for correct data input:

- Create and maintain a set of processes known as Input Editing Standards (IES). These are the rules for data elements that must be followed at the point of entry. For example, you might standardize all references to the International Business Machines corporation as IBM. You would require that 2-digit state abbreviations conform with USPS standards. And you would require that all titles be spelled out fully. Most companies create an input standards document when they first create a computerized database of customer information. But over the years that document may get lost, out-dated or filed somewhere collecting dust. Your first step is to find that document, review it, refresh it, and put it into use.
- Train data entry personnel on the IES rules, and repeat the training at least

quarterly. It's not just for new employees, but also needed as an ongoing refresher. A corollary point: Don't expect to pay your key-entry personnel peanuts and get great results. They need substantial training and incentives to do a good job maintaining your data asset.

- Use address-checking software at point of entry, to ensure deliverability.

Starting out with clean data is only the beginning. Business data tends to degrade at the rate of 3-6% per month, so you must invest in ongoing maintenance. Here are the best manual methods for data cleanliness:

- Train and motivate employees who have direct customer contact to request updates at each encounter. This includes call center personnel, customer service, sales people and distributors. It may be the job of marketing to keep the database clean, but data is a valuable corporate asset, and everyone has a stake in its quality.
- Segment your file, and conduct outbound confirmation contacts for the highest value accounts. This can be by mail, email or telephone.
- When using first-class mail, request the address correction service provided by the USPS. Put in place a process to update the addresses from the "nixies," meaning the undeliverable mail that is returned to you.
- Invite your customers to help you maintain their information correctly. Make the contact information available on a password-protected website, and ask your customers to key-enter changes as they occur. Offering them a good reason to do so, or perhaps a premium or incentive, will result in higher levels of customer compliance.

## THE AUTOMATED METHOD

Once you have manual methods underway, send your data out to a service provider for regular clean up. We recommend data cleansing at a third party at least twice a year. Large providers of business data are skilled at matching your file to their databases of standardized, updated records, and giving you back the good information. On a per-record basis, automated clean-up is inexpensive, and should be combined with an ongoing manual program of data hygiene.

There is quite a bit of misapprehension about the nature of automated data hygiene. Because it involves a matching process against a larger national database, some people confuse it with other data processes. So, before we go into more detail about what it can do, let us be clear about what we don't mean by automated data clean up for business marketers.

Sending your names out for clean up is not to be confused with:

- De-duplication, which means identifying records that qualify as duplicates.
- Data append, which means adding extra fields like an industry code, years in business, a credit score, or company size.

At the same time, it's important to realize that automated data hygiene cannot clean up everything on your database. For example, changes to a person's title or direct phone number are unlikely to be reported with any speed into a national database. So much of the time, the vendor will have no fresher title or phone data than you have yourself. And there's another matter to consider: Whose data is correct? If the name you have on your file for a company or a person is different from the name on the national database, how will you decide which one to accept? Most companies give preference to the data that was most recently collected or confirmed with customers.

The national databases maintained by various vendors have only one ultimate standard against which address accuracy can be measured, namely, the USPS. In fact, the only "true" addresses, street, state and ZIP code, are those recognized by the post office. So you can count on the outside vendors to clean up addresses to the point where they will support mail delivery. But you won't have the same level of confidence in the potential clean-up of telephone numbers, fax numbers, email addresses, and job titles. For such elements, verification via outbound contact and/or inbound web-based updating is the only method to ensure accuracy and timeliness.

This may be a disappointment to those who were hoping that they could simply "send our data out for clean up." In fact, the best method for ongoing maintenance of many important data elements used by business marketers is outbound contact and verification. Because this is an expensive and time-consuming process, we recommend that you verify your most valuable accounts first, and then decide the benefit of continuing on to your lesser value accounts.

For a thorough understanding of what outside vendors can and cannot do for you via automated hygiene processes, we conducted a research study in 2004 that involved clean up of a sample file by four leading suppliers who have deep experience with business data. Please review the research results presented in this report.

## THE DATA HYGIENE COMPARATIVE ANALYSIS PROJECT

In the spring of 2004, we invited a group of leading business-to-business data services providers to join us in a research project to compare their various approaches to data clean up. Four vendors agreed to participate: Acxiom, DataFlux, Donnelley Marketing and Harte-Hanks. We compiled a sample file of 10,000 “live” names, containing 12 fields. The names we used came from a variety of client sources. They were all names of individuals at business addresses.

We asked the vendors to perform their typical hygiene processes on the data and send the results back to us within 30 days. We also asked them to answer some questions about their companies and their approaches to data hygiene. Finally, we requested that they do this work at no charge, for the benefit of members of the business marketing community.

As you can imagine, this research project raised several fairly touchy issues. First, we were asking the vendors to open their doors, and reveal the results of their processes as compared to their direct competition. We gratefully acknowledge the courage and openness of the vendors who chose to participate. To reduce the competitive pressure, we are withholding the identity of the specific vendors in the report below (Table 5), which reveals the number of data elements corrected, by vendor.

Second, everyone involved in the project recognized the importance of protecting the privacy of the business people and companies whose names happened to turn up on the sample file. If the vendors were to apply their actual standard data hygiene processes to the file, live data was required. To protect the privacy of those involved, we have decided not to publish the sample records of individual names and addresses after clean up.

### TABLE 1: VENDOR DEFINITIONS OF DATA HYGIENE

*To make sure we were all talking about the same thing, we asked the vendors, “What is your company’s definition of data hygiene?”*

#### **Acxiom**

“Purpose-driven data management practices and/or processes that promote data accuracy. Typically is applied to name and address data content, correction and completion.”

#### **DataFlux**

“A 5-phase data management cycle, including profiling (inspection), quality (correction), integration (merging and linking), augmentation (enhancement), and monitoring (auditing and control). Understand the data problems: improve the data.”

#### **Donnelley Marketing**

“A broad range of processes that collectively deliver the highest deliverability of an address: standardizing, correcting, updating and verifying.”

#### **Harte-Hanks**

“The process of solving business problems resulting from inadequate data quality: accuracy, completeness, timeliness, validity.”

**TABLE 2:  
VENDOR-DESCRIBED DIFFERENTIATION**

*We thought readers would find it helpful to understand how the vendors view themselves in comparison with their competition. So we asked the vendors, “What are the 5 most important ways your work differs from your competitors?”*

**Acxiom**

1. Ability to recognize and parse name, business name, and address components
2. Ability to recognize the difference between business and consumer entities
3. Ad hoc, batch, automated batch and real time support of hygiene solution delivery
4. Abilitec-enabled occupancy database tool
5. NCOA/ChangePlus

**DataFlux**

1. Integrated data profiling and data quality technologies
2. Technologies developed in house, using the same core engine
3. Forthcoming capability to monitor data quality over time
4. Interface permits business users (non IT staff)
5. All processing done in one pass

**Donnelley Marketing**

1. Proprietary file of 13 million businesses in the US and Canada, and over 8 million executive contacts and title elements, for verification, appending and addition of missing elements
2. 100% telephone verification of each business record at least once a year
3. Ability to track executives at their home or business addresses
4. OnePass system allows all B-to-B hygiene to be done in one continuous logical flow
5. Proprietary Mailability Score that ranks each address based on deliverability

**Harte-Hanks**

1. Expertise in multiple vertical markets
2. Ability to provide customizable and flexible client-specific hierarchy of business rules
3. Objective selection of best-of-breed vendors of business files, to suit client needs
4. Integrated access to both USPS advanced postal products and HH proprietary data
5. Broad data management tool set

**TABLE 3:  
VENDOR DEFINITIONS OF “BAD” DATA**

*To understand any potential differences in the subject matter, we asked,  
“How do you characterize ‘bad’ B-to-B data?”*

**Acxiom**

“Data that fails to meet specific data content requirements and/or cannot be used to fulfill a specific business purpose.”

**DataFlux**

“Any data that does not support the underlying processes or business applications built on that information.”

**Donnelley Marketing**

“There is no right answer. The answer is to look within the customer segment and identify what is ‘bad’ to them.”

**Harte-Hanks**

“Data that fails to support the mission of delivering the right message to the right individual through the right channel.”

**TABLE 4:  
THE SAMPLE FILE**

*Our compiled file of business names and addresses came from a variety of client sources.  
We aimed for 10,000 names, and resulted in 9,699 usable records.*

**Each record contained the following fields:**

Last name  
First name  
Phone number  
Fax number  
Email address  
Business title  
Company name  
Address 1  
Address 2  
City  
State  
ZIP code

**TABLE 5:  
TOPLINE CORRECTION COUNTS ON THE 9,699 RECORDS**

*Each vendor reported slightly different counts, based on match rates and the data the vendors have on hand. You will notice that there are wide fluctuations in the counts on non-postal data, like fax, phone and email. This is because some vendors own current data, while others rent or lease it from third parties as needed by their clients. For this research project, we did not want the vendors to incur any out-of-pocket expense, so data owners delivered higher counts in these categories. Another reason for discrepancies is the way the vendors defined certain fields, like Address line 1 versus Address line 2. What strikes us as we look at these results is, for the most part, how similar they are.*

	Vendor 1	Vendor 2	Vendor 3	Vendor 4
ZIP codes corrected	344	479	446	864
ZIP+4s added	8652	9375	8706	9101
Carrier routes coded	9342	9381	9401	9121
Delivery points coded	9333	9324	9203	9101
Street addresses corrected (addr1)	1776	5387	7583	685
Street addresses corrected (addr2)	704	1588	NA	3007
City names corrected	592	472	470	648
State codes corrected	163	472	470	158
Phone numbers appended	2101	4097	NA	4183
Fax numbers appended	0	2931	NA	2834
Email addresses appended	520	861	NA	645
NCOA matches	760	760	778	689
Delivery point validated records	9151	9212	9203	9086
CASS certified records	9645	9596	9334	9589

## CONCLUSIONS & OBSERVATIONS

This study suggests that large, reputable vendors will provide very similar services when it comes to postal address standardization and correction, such as ZIP+4 and NCOA (National Change of Address). The marketer should not ask the vendor to change a title or company name — this information must come from the customer. Marketers can request that the vendor provide

phone numbers, fax numbers and email addresses, but these are not part of standard data clean up as defined by most vendors. The append rates for these elements will differ by vendor, and no vendor can provide 100% coverage. In short, it's the marketer who must make the final call about customer data accuracy.





**BERNICE GROSSMAN** is president of DMRS Group, Inc., a marketing database consultancy in New York City. She is past chair of the B-to-B Council of The DMA. Reach her at [bgrossman@dmrsgroup.com](mailto:bgrossman@dmrsgroup.com)



**RUTH P. STEVENS** consults on customer acquisition & retention, and teaches marketing to graduate students at Columbia Business School. She is the author of *The DMA Lead Generation Handbook*, and her new book is *Trade Show and Event Marketing*, now available at Amazon. Reach her at [ruth@ruthstevens.com](mailto:ruth@ruthstevens.com)

---

This publication is part of a series entitled *Business-to-Business Database Marketing*, by Bernice Grossman and Ruth P. Stevens. Papers published to date include:

“Our Data is a Mess! How to Clean Up Your Marketing Database”  
(October 2005)

“Keep it Clean: Address Standardization Data Maintenance for Business Marketers”  
(February 2006)

“Outsourcing Your Marketing Database: A ‘Request for Information’ is the First Step”  
(March 2006)

These papers are available for download at  
[www.dmrsgroup.com](http://www.dmrsgroup.com) and [www.ruthstevens.com](http://www.ruthstevens.com)

---