ENHANCING YOUR B-TO-B DATABASE WITH DATA APPEND

BY BERNICE GROSSMAN & RUTH P. STEVENS
JANUARY 2007

ENHANCING YOUR B-TO-B DATABASE WITH DATA APPEND

EXECUTIVE SUMMARY

Database append provides a convenient and cost-effective way to enrich the customer information in your marketing database. Many database marketing vendors offer append services, and business marketers are well advised to explore the opportunity to enhance the accuracy and predictive power of their data. This paper explains the process of data append, and reveals the results when three leading vendors appended data to a particular sample of 10,000 business records.

Business marketers can enhance their customer information by appending certain data elements purchased from third-party data vendors. Compared to collecting the data directly from customers and prospects, data append is a fast and convenient way to enrich your database, for purposes of research, analysis, modeling, and campaign selection.

To explore the ins and outs of data append, here is a step-by-step review of the append process. We also share the results of a piece of original research that is intended to illustrate the append process and to demonstrate the differences among append vendors.

GETTING STARTED IN DATA APPEND

Let us begin with a definition of data append:

> Enhancing your customer and prospect file with additional data fields (also known as "firmographics") supplied by commercial data compilers.

The data append process in B-to-B is no different from that performed on consumer files. Before you begin, you must decide on the fields you want to append. While appended data is cheaper than proactively collected data, there is still a cost associated with it. So only buy the data elements that will drive measurable value for your firm.

Most B-to-B marketers find the greatest value from knowing the size and industry of their customer and prospect companies. But there are scores of business data elements available for purchase, among them:

- Actual Number of Businesses at Multi-Tenant Code
- Affluent Neighborhood Location Indicator
- Block Group
- Business Status Code (Headquarters, Branch, Subsidiary)
- Census Tract
- City Population Size
- Contact Name, Title, Salutation, Gender, Ethnicity
- Credit Rating
- Entrepreneur Indicator
- Fax Number
- Female Executive/Owner Indicator
- Foreign Parent Indicator
- Fortune Magazine Ranking
- Geocode
- Government Segment Code (Federal, State, County, Municipal)
- Growing/Shrinking Indicator
- High Income Executive Indicator
- High-tech Business Indicator
- Inport/Export Code
- Location Property Manager

- MSA (Metropolitan Statistical Area) Code and Description
- NAICS and Description
- New Business Code
- Number of Employees (Site and Total Company)
- Number of Personal Computers
- Office Size (Employees, Square Footage)
- Own/Lease Code
- Phone Number
- Population Density
- Post Office Box
- Public/Private Indicator
- Public Filing Indicator (Bankruptcy, Lien, Foreclosure)
- Sales Volume (Site and Total Company)
- SIC and Description
- Size of Yellow Pages Ad
- SOHO Business Indicator
- Stock Exchange Ticker Symbol
- Toll-free Number
- Web Site URL
- White Collar Indicator
- Year Established

Many of the kinds of data that B-to-B marketers append to their files are not particularly fast-changing data elements. So, the frequency of appending or updating appended data is a function of how often you update your marketing database, the cost/benefit of updating, and the uses to which the data will be put. For example, if you are profiling your customer base in order to develop your marketing strategy, you will be able to work with data that hasn't been refreshed in several months. On the other hand, if you are selecting campaign targets, you will want to refresh your appends regularly, for attachment to new records that have come onto your file.

HOW TO APPEND DATA, STEP BY STEP

Step 1: Clean up your file

Your append efforts will be much improved if you begin with clean data. Inaccurate data is harder to match, and will result in lower append rates. To clean up your file, perform the following standard hygiene steps:

a) Run your own standard in-house hygiene protocols.

b) Ask your vendor to run their hygiene processes, like NCOA (national change of address), correction of state abbreviations, ZIP Codes and ZIP+4, append of directionals, etc.

c) Review your data. There may be anomalies that are identifiable to the naked eye that would be missed — or mangled — by a computerized process. For example, you may know that Avenue Magazine would not appreciate being addressed as Ave Magazine, which is how an automated hygiene process might standardize it.

Step 2: Create a list of vendor candidates

There are several large and mid-sized data vendors with B-to-B expertise. You need to identify the one that will perform most effectively with your data.

To find the right vendor, conduct the following steps:

a) Create a short (3-5) list of vendors. Make sure they have B-to-B experience. Some of the leading vendors in this space are: D&B, Equifax, Experian, InfoUSA and MarketModels. But there are many others.

b) Ask them for a list of fields they can append, since there will be differences, by vendor. Make sure your chosen fields are available.

c) Ask how recently their data has been collected, by field, and the source of their data.

d) Ask them to show you a sample of the results of their matching process, so you can get a feel for its flexibility and appropriateness to your data.

Step 3: Perform a test of several vendors

The effectiveness of any given vendor is a function of how the vendor performs on your data, not on any one else's files. So the only way to identify the best vendor for you is to conduct a live test. Here is the process:

a) Explain to the candidate vendors that you want them to run their append process on a sample of 10,000 names from your file. Depending on the size of your file, you may want to select a somewhat smaller or larger sample.

b) Seed the file with some companies about which you already have detailed — and correct — information. This will allow you to do a quality check on the returned append input.

c) Ask the vendors not to run any data hygiene or de-duplication for this test. This will allow a better apples-to-apples comparison of the data append process on your records.

d) Give the participating vendors several weeks to do the work, and ask them not to charge you for the test. However, do ask them for a bid on the whole project so they know that you intend to pay for the service once the project is awarded.

Step 4: Determine your evaluation criteria

How will you recognize success when you see it? There are four criteria by which to judge the relative performance of the vendors:

1. Match rate. Defined as the number of your records identified as also appearing in the vendor's database, divided by the number of records the vendor received from you. Match rates of 50% are not uncommon in B-to-B data environments.

2. Hit rate. Defined as the number of matched records that also had the required fields. available for append, divided by the number of records the vendor was able to match against the vendor's database. Hit rates range widely, depending on the data element required. You may experience a 100% hit rate for a easy-to-find data element like SIC, while the more esoteric elements like Size of Yellow Pages Ad will be much lower.

3. Accuracy. Expect a certain amount of incorrect data to be part of the append process. It's inevitable. The only way to measure the accuracy of the appended elements is by seeding your file with names of companies about which you already know the correct answers, and compare the results.

4. Price. We have made price the last criterion for a good reason. Assuming you are purchasing data that will drive real business value for your company, the amount of correct data you can expect to get will be the most important factor. However, prices can vary dramatically, and you may have budget constraints, so price is a factor you need to consider.

Step 5: Assess your test results

Based on the vendors' performance against your selection criteria, your selection decision should be fairly straightforward, as a combination of performance and price. Some points to keep in mind as you make your decision:

a) Not all vendors have the same data, despite what they may tell you. Similarly, not all vendors collect data is the same way and from the same sources.

b) Don't be surprised if less than half of your file will be matched. In short, B-to-B data is a bear.

c) Beyond that, only a percentage of the matched files will be appendable.

d) Your files will have their own unique hit rates, which is why you need to conduct the experiment.

e) You may not be able to get every data field you need. While third-party data append is the lowest-cost source of business data, it's only the first step in a longer process of making your database as useful as it can be.

THE DATA APPEND COMPARATIVE ANALYSIS PROJECT

In the spring of 2006, 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 append. Three vendors agreed to participate: InfoUSA Donnelley Marketing Division, Equifax and MarketModels

We compiled a sample file of 10,000 "live" records, from a variety of client sources. We expected that they were all records of individuals at business addresses, although we later learned that a sizable number of consumer addresses had crept in. The records comprised 9 fields:

- Last name
- Address 2
- First name
- City
- Business title
- State
- Company name
- ZIP code
- Address 1

We asked the vendors to perform their typical append processes on the data, but not to perform any data hygiene or de-duplication on the file we sent. Our reasoning was that this is an append project. We did not want to introduce an additional variable in the form of data hygiene services that would impact the result.

We asked the vendors to take no more than 14 weeks to do the project (usually 3 weeks is sufficient). Finally, we requested that they do this work at no charge, for the benefit of members of the business marketing community. Our sincere thanks to them all for participating.

Our instructions were to append the following fields, which are very typical of B-to-B append projects:

- Year Started: When the company went into business
- SIC: Although NAICS is available, many users and vendors still rely on SIC to identify the company's industry category. We asked for primary SIC
- Franchise: Indicates a franchise business like Jiffy Lube
- Headquarters/Site: Does the address reflect a company headquarters or simply a site, meaning, a non-headquarters location
- Number of Employees Total: Number company-wide
- Number of Employees Local: Number at this site
- Sales Volume Total: Sales company-wide
- Sales Volume Local: Sales at this site
- Public/Private
- Minority Owned
- Government
- Educational Institution

RESULTS

As you review the append project results, you'll notice that some data fields were not provided by some vendors. This may be because:

- The data was not available to this service provider.
- The service provider chose not to offer this data element for this project

Now, let us examine the data as appended by the three participating vendors to the sample file they received.

TABLE 1:
AVERAGE APPEND RATES BY FIELD

The highest level of appendability was achieved for Number of Employees Local, SIC, and Sales Volume Local. The other fields experienced statistically insignificant append rates. Keep in mind, however, that these results are specific to this particular sample data set, so they cannot serve as any sort of benchmark.

	SIC	Head-quarters/ Site	Number of Employees Local	Sales Volume Local	Public/ Private
Average/ Received	50.49%	28.95%	50.37%	44.92%	21.53%
Average/ Matched	97.83%	56.69%	98.03%	89.22%	44.44%

TABLE 2:
MATCH RATES

All three vendors identified 9,665 records received. They then attempted to match these records to their own databases, with widely varying results, as seen in this table. Keep in mind that the relative match rate is primarily a function of how the matching process of each vendor interacts with the data sample submitted. On another sample, the relative match rates are likely to be very different. This table illustrates the importance of testing each vendor with your own data sample.

Vendor	Records Matched	Match Rate
InfoUSA	5,982	61.89%
Equifax	4,904	50.74%
MarketModels	4,074	42.15%

When considering the results of matching your data to the vendor's database, there are a number of drivers that may explain how your data performed:

- Some of your records may have contained residential address data. You may want to ask the vendor for residential address counts. In this test, InfoUSA Donnelley Marketing Division voluntarily provided a count of consumer records on our file (see Table 15).

- Company names can be written so many ways — IBM versus International Business Machines, for example. If your files are not standardized, you will experience lower match rates.

- Companies may use many addresses, including post office boxes.

- You may have more duplicates than you thought.

- Your data needs to be cleaned up.

TABLE 3:
YEAR STARTED

	Hits	Hits/Received	Hits/Matched
InfoUSA	1,928	19.95%	32.23%
Equifax	2,711	28.05%	55.28%
MarketModels	Not provided	Not provided	Not provided

TABLE 4:
YEAR STARTED

	Hits	Hits/Received	Hits/Matched
Donnelley	5,982	61.89%	100%
Equifax	4,585	47.44%	93.50%
MarketModels	4,074	42.15%	100%

TABLE 5:
FRANCHISE

	Hits	Hits/Received	Hits/Matched
InfoUSA	560	5.79%	9.36%
Equifax	Not provided	Not provided	Not provided
MarketModels	28	0.29%	0.69%

TABLE 6:
HEADQUARTERS/SITE

	Hits	Hits/Received	Hits/Matched
Donnelley	2,033	21.03%	33.99%
Equifax	2,654	27.46%	54.12%
MarketModels	3,708	38.34%	90.97%

TABLE 7:
NUMBER OF EMPLOYEES, TOTAL COMPANY

	Hits	Hits/Received	Hits/Matched
InfoUSA	800	8.28%	13.37%
Equifax	Not provided	Not provided	Not provided
MarketModels	Not provided	Not provided	Not provided

TABLE 8:
NUMBER OF EMPLOYEES, LOCAL (AT THE SITE)

	Hits	Hits/Received	Hits/Matched
Donnelley	5,628	58.23%	94.08%
Equifax	4,904	50.74%	100%
MarketModels	4,074	42.15%	100%

TABLE 9:
SALES VOLUME, TOTAL COMPANY

	Hits	Hits/Received	Hits/Matched
InfoUSA	697	7.21%	11.65%
Equifax	Not provided	Not provided	Not provided
MarketModels	Not provided	Not provided	Not provided

TABLE 10:
SALES VOLUME, LOCAL (AT THE SITE)

Notice that Equifax and MarketModels identified the same number of hits for both Sales Volume Local and Number of Employees Local.

	Hits	Hits/Received	Hits/Matched
Donnelley	4,047	41.87%	67.65%
Equifax	4,904	50.74%	100%
MarketModels	4,074	42.15%	100%

TABLE 11:
PUBLIC/PRIVATE

The following four tables are structured on a "yes/no" basis, so a hit means that the vendor had information on whether the company is public or private, minority owned, etc.

	Hits	Hits/Received	Hits/Matched
InfoUSA	1,338	13.84%	22.37%
Equifax	2,281	23.60%	46.51%
MarketModels	2,625	27.16%	64.43%

TABLE 12:
MINORITY OWNED

The determination of minority owned business here is driven by the federal government's very detailed set of characteristics, and may not be entirely consistent with how a marketer would define it.

	Hits	Hits/Received	Hits/Matched
Donnelley	127	1.31%	2.12%
Equifax	129	1.33%	2.63%
MarketModels	534	5.53%	13.11%

TABLE 13:
GOVERNMENT

	Hits	Hits/Received	Hits/Matched
InfoUSA	343	3.55%	5.73%
Equifax	362	3.75%	7.38%
MarketModels	47	0.49%	1.15%

TABLE 14:
EDUCATION

This particular file did not happen to contain many educational institutions. But there is one other technique you can use to assess this factor: Compare the Education hit rate against the education SIC codes appended to the file.

	Hits	Hits/Received	Hits/Matched
Donnelley	357	3.69%	5.97%
Equifax	139	1.44%	2.83%
MarketModels	140	1.45%	3.44%

TABLE 15:
CONSUMER ADDRESSES

While we did not request information about consumer addresses, InfoUSA Donnelley Marketing Division supplied an analysis based on their consumer databases, so we include it here. The relatively high level of consumer addresses on this file is certainly an impediment to perfect match rates. However, it is very common in B-to-B, due to the preponderance of small office / home office (SOHO) businesses, as well as situations where business customers respond to marketers using their home addresses. If you do find a sizable set of consumer addresses, you may want to conduct additional protocols on that portion of the data.

	Hits	Hits/Received	Hits/Matched
InfoUSA	1,370	14.17%	22.90%
Equifax	Not requested	Not requested	Not requested
MarketModels	Not requested	Not requested	Not requested

OBSERVATIONS & CONCLUSIONS

This study suggests that large, reputable vendors will provide very similar services when it comes to data append. Results are heavily a function of the nature of the names processed, so it is imperative that you test a sample of your file before selecting a vendor. We did not ask the vendors to clean up the data, so these match-rate results are on the low side of what's possible. The highest hit rates on this particular data set came from SIC, Number of Employees Local and Sales Volume Local. The lowest hit rates were from Government, Education and Franchise. The only ways to measure append accuracy are 1) seed the file with names you already know the truth about, and/or 2) conduct outbound telephone verification on a sample.

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

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

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)

"15 Thorny Data Problem That Vex B-to-B Marketers, and How to Solve Them"
(January 2007)

These papers are available for download at
www.dmrsgroup.com and www.ruthstevens.com