B2B Contact Data Quality Maintenance:
Onsite or Outsourced?
Solution Pros and Cons

OAUG: COLLABORATE08
April 16, 2008

Andrew Kapochunas
Leader, Customer Data Quality
Dun & Bradstreet
kapochunasa@dnb.com

<u>Every</u> MDM initiative involves the selection & implementation of solutions to correct & maintain the quality of basic B2B contact information:
– business name
– mailing/physical address
– contact name
– telephone number
– email address

A wide array of "onsite licensable" and outsourced solutions are available: choosing <u>between</u> them, or choosing the right <u>combination</u>, can be daunting. This paper outlines the basic choices, with pros and cons of each as partial & total solutions.

The Basic Problem: Business Data Decay

Percentage of data that has changed

Age of Data	CEO	Other Execs	Physical Address	Phone Number	Company Name
3 months	5.3%- 11.4%	5.4%- 13.6%	5.2%- 6.8%	4.5%- 5.7%	3.1%- 9.1%
6 months	10.6%- 22.8%	10.8%- 27.1%	10.4%- 13.7%	9.0%- 11.4%	6.2%- 18.2%
9 months	15.9%- 34.1%	17.9%- 40.7%	15.5%- 20.5%	13.5%- 17.0%	9.3%- 27.3%
1 year	21.2%- 45.5%	21.5%- 54.2%	20.7%- 27.3%	18.0%- 22.7%	12.4%- 36.4%

<u>Low end of change range: All US businesses</u> on the Dun & Bradstreet database of 44 million US businesses. <u>High end: businesses with "Year Started" 24 months or less</u>: 11% of all US businesses

Solving contact data decay of:

1. **Address**
2. **Business Name**
3. **Contact Name**
4. **Telephone**
5. **email**

A complete solution for any or all elements must include addressing multiple factors:

a. Address Accuracy

"Why should I care about address accuracy? I'm not building my database to support direct mail"

If you care about accurately matching your customer data between internal databases to understand duplication – and gain a 360-degree view...

If you care about accurately matching 3rd-party data to append firmographics, credit and email information...

Then you should care about address accuracy

All USPS software and metrics are not the same:

1. **Second Generation Delivery Sequence File (DSF2)**
   1. Corrects/scores/categorizes per USPS's largest address database
   2. It's available batch Outsourced, only

2. **CASS-certified Address Standardization, Address Correction and ZIP+4 appending**
   1. Corrects according to a 1983-era metric, with a reduced address database: address ranges, as well as actual addresses
   2. DPV option gives the DSF2 "view" of an address corrected via Address Standardization software: "Y," "S," "D," and "N"
   3. Available for transactional or batch, Outsourced or Onsite license, alone, or bundled with ETL/matching software: Trillium, Firstlogic, etc.

But is it just a matter of comparing costs in deciding on an address accuracy solution? NO. Results depend on the expertise of the person or vendor applying the software. Results are not generic.

Ineptly configured/implemented onsite software will negatively affect your data quality: TEST the effect of adding it to your work flow with "before/after" benchmarking against an outsourced vendor.

Results by vendor can vary by 12% or more, depending on expertise in parsing business addresses.

Case History # 1

2005: match of client data to D&B data, by D&B:

Total raw client input:	**589,180**
Records having street/PO Box address:	**491,462** *(83.4%)*
Total client-accepted matches:	**430,334**

73% match to raw data, 88% match to data with addresses

2007: client purchase & implementation of onsite address standardization software:

Total "improved" client input:	**357,170**
Records having street/PO Box address:	**239,411** *(67.0%)*
Total client-accepted matches:	**28,893**

8% match to *"improved"* data, 12% match to data with addresses

---

b. Address Completeness

An address can be "Accurate," yet incomplete:

- **DSF = 1/DPV = Y = 100% accurate to the suite #, *and deliverable via Standard or First Class mail as long as the company is still there, and still using the same name***

- **DSF = 3/DPV = D = building number is accurate, but location requires a secondary address for mail delivery, you're missing it. *Less likely to be delivered by First Class, undeliverable via Standard Class mail.***

- **DSF = 8/DPV = N = building number missing. *Ability for D&B to get a confident match to such a record is reduced by 75%***

Solving for address completeness:
1. CASS-certified Address Standardization, Address Correction and ZIP+4 appending: SuiteLink option: adds/corrects suite/floor numbers according to a carrier-compiled database of business names and suite numbers
    1. Available for Onsite license/Outsourced processing

2. Match, either before or after postal processing, to a compiled business database to add/correct building and suite numbers, when you know that yours are wrong or missing, and the ones on the compiled database are right
    1. *D&B has a patent-pending automated outsourced solution*

a. Address Timeliness

An "accurate" address for a customer is useless if your customer moved from it two years ago.

1. Address Standardization – *or any address correction software or database* – onsite or outsourced, cannot track a business when it moves.

2. By and large, your customers will <u>not</u> tell you when they move.

Case History # 2:

End-user-applied Address-Standardized-only file benchmarked via an outsourced multi-step process that includes full NCOA:

Benchmark	%	Compared with Avg. B2B Customer File Processed
ACCURACY		
Verified "accurate" addresses on input:	90.8%	55% better than average
Addresses at 100% accuracy:	82.5%	19% more than average
RECENCY		
Customers not at submitted address:	16.3%	3.6x worse than average
Additional Customers possibly moved:	3.0%	3.8x worse than average
Definite + possible stale addresses:	19.3%	

A very accurate file maintained by a high-tech firm unaware that nearly a fifth of its customers were at a an address different from the one they had verified as "accurate"

Solutions for Address Timeliness:

1. **USPS NCOALink**: two versions – <u>batch processing, only</u>
   a. **"Full" = <u>48 months</u> of moves; <u>updated weekly</u>; <u>use your own matching software</u>**

      1. **Licensable by vendors, only – an end-user would have to make the process available to 3rd-parties**

   b. **"Limited" = <u>18 months</u> of moves; <u>updated monthly</u>; <u>USPS matching software</u> (= lower hit rates)**
      1. **Licensable by end-users and by vendors**

2. **USPS NCOALink Concerns**

   a. *USPS did not build NCOALink to help you maintain your customer database – they built it to be used as a last-minute address update for mail addressed to that business. They are interested in where a business wants its mail to be delivered, <u>not </u>where the business is physically located*

   b. *Leverage NCOALink, and you will get "moves" that are physical-to-P.O. Box – the location where the business wants mail delivered, which could be headquarters box, or a box in another state where all branch/subsidiary mail is to be sent to facilitate efficient processing*

_____

Case History # 3:

End-user-applied Address-Standardized file sent to outsourced "full" NCOA licensee, vs. same file outsourced to vendor for application of multi-step postal correction/updating, including DSF2 and "full" NCOA:

Input: 100,000 U.S. business records

	Onsite + Licensee "X"	100% Outsourced
NCOA hits with new address:	2,495 2.5%	5,448 5.4%
NCOA hits *without* new addresses:	<u>23</u>	<u>283</u>
	2,528 2.5%	5,731 5.7%

"Lift" in NCOA hit rate due to superior business record parsing, use of 100% outsourced multi-step solution (Address Standardization/ DSF2/LACSLink/NCOALink):   127%

_____

Solutions for Address Timeliness, continued

3. LACSLink: formerly stand-alone, now bundled with Address Standardization software
    a. 6 million-record crosswalk of radically changed addresses, themselves
    b. Rural Route to building number and street changes
    c. Street name changes

4. Matching to a privately compiled/maintained database of moves to get new addresses for businesses:
    a. Who religiously tells a company that they have moved? Subscribers to trade magazines. And, what, precisely, are they saying – that their company has moved, or that they want to get their subscription delivered to another address – *perhaps at home?* When you "move" your business subscription to your home address, are you careful to remove your company name from the new address?
    b. USPS move types: Individual, Family and Business

5. Matching to a privately compiled/maintained database of businesses that includes historical, or "former" business addresses might enable identification of true business moves incremental to those on NCOALink
    a. *Up to 30% of business moves today not filed with USPS*
    b. At least two business compilers, Dun & Bradstreet and Experian, claim to have "former" addresses for businesses, and so might be able to provide incremental updates to addresses even after NCOALink processing:

    Addresses on D&B's US Match File

Match points for US NextGen (1007):		
Physical Addresses	51,927,911	
Mail Addresses	6,779,941	
Total AOS Match points		58,707,852
Match points for HISTORICAL Data:		
Physical Addresses	2,319,368	
Mail Addresses	116,070	
Total HISTORICAL Match points		2,435,438
Match points for Executive Home Address Data:		
Physical Addresses	9,101,174	
Mail Addresses	258,652	
Total EHA Match points		9,359,826
Total match reference points :		70,503,116

Leveraging address updates from a compiled business file:

Matched Elements	Count	% Count	MDP Field Notation
Building No. & Street Matches	261,710	100	Digits 3-4, 5-6
Current Building Number and Street	246,575	94.22	xx-00-00-xx-xx-xx-xx
Former Building Number and Street	5,753	2.20	xx-02-02-xx-xx-xx-xx
Home Building Number and Street	9,382	3.58	xx-03-03-xx-xx-xx-xx
P.O. Box Matches	11,257	100	Digits 11-12
Current Address	11,112	98.71	xx-xx-xx-xx-xx-00-xx
Former Mail Address	66	0.59	xx-xx-xx-xx-xx-02-xx
Personal Mail Address	79	0.70	xx-xx-xx-xx-xx-03-xx

Match results on a file that already generated new addresses from a multi-step postal process that included matching to NCOALink. Incremental matches allowed the customer to programmatically update additional customer addresses

d. Address Relevance

All your due diligence to maintain accurate, complete and up-to-date customer addresses can be stymied if you have irrelevant/inappropriate data in a field where both onsite and outsourced solutions expect to find a primary address. Such records are easy to find:
Alpha sort: 78,000 US "addresses":

ADDRESSL1	Count	Percentage
<blank>	2,811	3.56%
" METAL FABRICATORS, INC."	2	0.00%
" PRODUCTS, INC. "	1	0.00%
"10 EAST BACON ST,"	1	0.00%
"10675 ""E"" AVE - 8"	1	0.00%
"13-15 SHING WAN ROADTAI WAI, S	1	0.00%
"14 THORNE ROAD,	1	0.00%
"143 DORCHESTER ST #360,"	1	0.00%
"16-7, SHINDEN I-CHOME ADACHI-KU"	1	0.00%
"1750 112TH AV NE, STE C-100"	1	0.00%
"1777 CEBRIAN ST, RM SUPPLY"	1	0.00%
"1805 ENTERPRIZE DR, STOREROOM"	1	0.00%
"222 W LARCH ROAD, STE #A"	1	0.00%
"25 CHONG YIP ST KOWLOON , HON	1	0.00%
"2536 East Fender, Unit H"	1	0.00%
"266-310 HALL AVENUE E - RENFRE	1	0.00%
"2746 NE 45TH ST, SUITE 100"	1	0.00%
"2921 AVENUE B NORTH, BUILDING	1	0.00%
"30TH FLOOR"	1	0.00%
"3199 AIRPORT LOOP DRIVE, UNIT E"	1	0.00%
"33/10 MOO 4, CHAENG WATTANAT	1	0.00%
"334 NORWEST COURT, RM 108"	1	0.00%
"3455 HARVESTER ROAD, UNIT 22"	1	0.00%
"357 RUE FRANQUET ST-FOY, QC G1P	1	0.00%

Sort: repeated data, descending sequence:

ADDRESSL1	Count	Percentage
\<blank\>	2,811	3.56%
UNKNOWN ADDRESS	594	0.75%
NOT GIVEN	554	0.70%
CUSTOMER WILL CALL	73	0.09%
U S WEST COMMUNICATIONS	73	0.09%
TAXABLE	70	0.09%
13835 NW US HIGHWAY 441	69	0.09%
ATTN: ACCOUNTS PAYABLE	66	0.08%
ACCOUNTS PAYABLE	49	0.06%
WILL CALL	48	0.06%
TAX EXEMPT	34	0.04%
259 N RADNOR CHESTER RD	26	0.03%
ATTN ACCOUNTS PAYABLE	26	0.03%
SOLECTRON OPS	22	0.03%
*	21	0.03%
ATTN ACCTS PAYABLE	21	0.03%
EXEMPT	19	0.02%
SUITE 101	18	0.02%
101 N INDIANA AVE	17	0.02%
455 ALLEGHENY BLVD	16	0.02%
800 PHILLIPS ROAD	16	0.02%
CST	16	0.02%
DO NOT MAIL	15	0.02%
POSTMASTER	15	0.02%
.	14	0.02%

Don't outsource a solution to address relevance: no software reliably understands the distinction between appropriate and inappropriate street address data, apart from identifying blatant problems: punctuation, only, for example

Put "Cash Only" in your primary address field, and the most sophisticated postal processing will probably code it as an inaccurate address: a street name missing a building number

License ETL software capability *(Informatica, Trillium, etc.)* that lets YOU control the establishment of nonsense tables/rules specific to your data to identify problems, and which will allow you to easily scan the balance of a record to find where you put the appropriate data

2. Business Name Decay:

Percentage of data that has changed

Data age:	CEO	Other Execs	Physical Address	Phone Number	Business Name
3 months	5.3%- 11.4%	5.4%- 13.6%	5.2%- 6.8%	4.5%- 5.7%	3.1%- 9.1%
6 months	10.6%- 22.8%	10.8%- 27.1%	10.4%- 13.7%	9.0%- 11.4%	6.2%- 18.2%
9 months	15.9%- 34.1%	17.9%- 40.7%	15.5%- 20.5%	13.5%- 17.0%	9.3%- 27.3%
1 year	21.2%- 45.5%	21.5%- 54.2%	20.7%- 27.3%	18.0%- 22.7%	12.4%- 36.4%

**Low end of change range: All US businesses** on the Dun & Bradstreet database of 44 million US businesses. **High end: businesses with "Year Started" 24 months or less**: 11% of all US businesses

2. Solving Business Name decay

a. Accuracy, b. Completeness, c. Timeliness

Your safest bet: maintain the company name in the display the company itself uses:
   1. _If the company spells out "Corporation," keep it_

      a. _"Standardizing" company names to adhere to an arbitrary standard may be an IT "BDP," but it is a "WDP" in maintaining a relationship with a customer_

   2. _Be aware that many companies today use multiple valid names that represent the same company at the same location – they are not necessarily separate businesses_

Business names on D&B's US match file:

Match points for US NextGen (1007):		
Business Names	44,236,757	
Tradestyle 1 Names	7,282,831	
Tradestyle 2 Names	1,440,183	
Tradestyle 3 Names	265,868	
Tradestyle 4 Names	64,660	
Tradestyle 5 Names	17,544	
"CEO" names	28,008,146	
Total AOS Match points		81,315,989
Match points for HISTORICAL Data:		
Business Names	1,364,771	
"CEO" names	1,367,381	
Total HISTORICAL Match points		2,732,152
Match points for DDM Data:		
Executive names	10,498,966	
Total DDM Match points		10,498,966
Total match reference points :		**94,547,107**

To better identify and consolidate company records, D&B keeps up to six concurrent valid versions of the company name, former names, and current and former CEO names: because records come in that way!

Consider doing that yourself on your customer master

Business Name relevance:

BUSINESS_NAME	Count	Percentage
\<blank\>	1,673,526	72.70%
Herbalife	134	0.01%
Mary Kay	97	0.00%
none	97	0.00%
Monster.com	88	0.00%
test	82	0.00%
Data Entry	77	0.00%
Smith Inc	65	0.00%
Coastal Vacations	53	0.00%
home	48	0.00%
Big Value Website	46	0.00%
Home	46	0.00%
Avon	44	0.00%
ebay	41	0.00%
Data Entry Business	36	0.00%
First Baptist Church	35	0.00%
N/A	34	0.00%
Ebay	33	0.00%
My Dollar Shop	32	0.00%
None	32	0.00%
First Assembly of God	30	0.00%
Passion Parties	28	0.00%
data entry	28	0.00%
Mary Kay Cosmetics	27	0.00%
SMC	26	0.00%

3. <u>Contact Name</u> decay

a. Accuracy
b. Completeness
c. Timeliness
d. Relevance

The Sales & Marketing Institute says 75% of business cards change in some way every year. There is no equivalent to NCOALink for individuals at a <u>business</u> location; USPS delivers to business sites, not to individuals at business sites

	% changed	
Last verified:	**CEO**	**Other Execs**
3 months	**5.3%- 11.4%**	**5.4%- 13.6%**
6 months	**10.6%- 22.8%**	**10.8%- 27.1%**
9 months	**15.9%- 34.1%**	**17.9%- 40.7%**
1 year	**21.2%- 45.5%**	**21.5%- 54.2%**

There is no equivalent to NCOALink for individuals at a <u>business</u> location; USPS delivers to business sites, not to individuals at business sites

COLLABORATE08   April 16, 2008   Andrew Kapochunas: D&B
B2B Contact Data Quality Maintenance: Onsite or Outsourced?

4. Solving <u>Telephone</u> decay

   a. **Accuracy**
   b. **Completeness**
   c. **Timeliness**

Today, disconnected phone numbers can be reassigned to another company or individual within 10 days – seeing if the number still rings is irrelevant

Last verified:	Wrong Phone Number
3 months	4.5%- 5.7%
6 months	9.0%- 11.4%
9 months	13.5%- 17.0%
1 year	18.0%- 22.7%

4. Solving <u>Telephone</u> decay

Periodically matching to a compiled business file with phone numbers will tell you:
   1. The number you have is the same one they have
   2. The number you have is different
   3. You have a number they don't have
   4. They have a number you don't have
...but it won't verify your phone number.

Solution:
1. Periodically apply updates to area code assignments by subscribing to an updating service like Telcordia:
     http://www.trainfo.com

2. Match to a compiled business file with known former phone numbers -- if yours is an exact match, replace with their current one

Match points for US NextGen (1007):		
Phone Numbers	38,400,131	
Total AOS Match points		38,400,131
Match points for HISTORICAL Data:		
Phone Numbers	808,033	
Total HISTORICAL Match points		808,033
Total match reference points :		39,208,164

In your own data, look for repeated numbers in excess of repeated company names; 800 numbers; numbers entered with the leading "1," which may result in truncating the last valid number

PHONE	Count	Percentage
<blank>	88	0.06%
9	72	0.05%
8008004489 Adult line	56	0.04%
5555555555	37	0.02%
9999999999	32	0.02%
7329493000	21	0.01%
8476983000	20	0.01%
0	19	0.01%
0000000000	19	0.01%
5075555555	15	0.01%
1	12	0.01%
1111111111	12	0.01%
999	11	0.01%
4048435000	9	0.01%
5045882000	9	0.01%
8008004239	9	0.01%
8472599600	9	0.01%
8775574487	9	0.01%
9734903400	9	0.01%
999999999999	9	0.01%
6142282674	8	0.01%
6506887066	8	0.01%
8008008000 Adult line	8	0.01%
8472998000	8	0.01%

5. Solving <u>email</u> decay

Decay rate of individual email addresses mirrors that of contact names

a. **Internal/onsite solutions:**

    1. **Sending emails <u>and getting a reply</u> confirms validity**

    2. ***Getting a bounceback does NOT necessarily confirm decay:* you must be able to properly interpret:**

        a. **"Hard" bouncebacks are truly undeliverable**

        b. **"Soft" bouncebacks have multiple causes**
            1. **The emailbox was temporarily full**
            2. **The subject line included words that triggered suppression by the company**
            3. **The email included graphics that triggered suppression**

b. **Outsourced:**

    1. **Your digital agency, your email append company, your email campaign resource – may well be able to determine the reason for the bounceback, as can D&B's primary email partner, Outward Media:**

- ***Email addresses are flagged with indicators for promiscuous and dark domains to facilitate the process of weeding out false positive results.***

- ***Email addresses enter the domain verification stage, where our servers talk to recipient servers to verify the email domains are still active - also able to verify 60% of the email addresses.***

- ***Email addresses are suppressed against Client's opt-out file***

- ***A permission pass is sent on behalf of Client to allow recipients to opt-out prior to receiving marketing or advertising communications and to provide a final verification of the email address accuracy.***

- ***Undelivered email address is analyzed to identify false negatives: soft vs. hard bounce-back analysis involving 18 categories of bounce-receipts***

Solution Summary: <u>Onsite Only</u>

- ✓ **Pros**
    - ✓ <u>Security</u>: your data stays behind your firewall
    - ✓ <u>Accessibility</u>: you correct/update data <u>when</u> you want to
    - ✓ <u>Customizability</u>: you correct/update the <u>way</u> you prefer to
    - ✓ <u>Cost</u>: license once; the more you use it, the better the ROI

- ✓ **Cons**
    - ✓ **Data quality often compromised: some verifying data is simply not available for onsite license**
    - ✓ **Licensable versions of verifying/updating products may have less frequent update cycles – you may not be using the most up-to-date information**
    - ✓ **Application/implementation requires training, skill, maintenance; results will vary from seller's processing**

Solution Summary: <u>Outsourced Only</u>

- ✓ **Pros**
    - ✓ **Shop for best available data, repeatable processing, results and turnaround – test results more likely to bear out on rollout**
    - ✓ **No investment in software, maintenance – can easily change vendors if results are not as promised**
    - ✓ **Possibility of one- or two-stop shopping, instead of relying on multiple, unconnected, onsite applications**

- ✓ **Cons**
    - ✓ **Your data <u>does</u> leave the house**
    - ✓ **You are dependent on <u>when</u> a vendor is able to deliver**
    - ✓ **They will do it <u>their</u> way**

Recommendation:

Leverage Onsite + Outsourced

- ✓ **Address: License onsite capability for ETL, point-of-entry Address Standardization + DPV, and matching for internal use; outsource periodic robust correction and updating**

- ✓ **Company name, contact name, phone, email: Get it right onsite, *and get it in the right, dedicated, field,* using drop-downs, verifying repeats and field logic wherever possible; get an outsourced persistent key appended, in addition to your own, to as many records as possible, and use it to simplify getting adds/changes/deletes from data vendors; periodically outsource re-match/re-verify to ever-expanding, ever-changing B2B data**

#################