

DSS 2000 EMPLOYMENT DATA DEVELOPMENT

The DSS point gis data for 2000 was developed using data obtained from a commercial data provider (infoUSA) and data from the VT Dept of Employment and Training (DET). The Vermont DET database is created through the state unemployment insurance program, which is commonly referred to as the ES-202 data. DET data is only available for jobs that are “covered” by the unemployment insurance program. Student fellowships, military, farm workers, and a few other assorted types of employment are not included in the ES-202 data. The ES-202 data is also covered by a confidentiality agreement with the state, which limits CCMPO’s ability to share this data.

The infoUSA data is created through a phone survey of the business whitepages listings. This slightly increased the coverage of employment to establishments not included in the ES-202 data. For this reason, and the lack of confidentiality restrictions, we decided to use the infoUSA data as the base for the DSS employment. We did supplement the infoUSA data with DET data for particular employers which were not included in the infoUSA data. The “Emp_type” field in the employment shapefile specifies the origin of the employee count for a particular employer. In some cases, there was no figure in the infoUSA data (the value of employee count was -9999), and so an estimate was made. There were also several employers recorded in the infoUSA data more than once, and these are listed as “MPO DUPLICATE” with a Dss_emp (the employee count) of zero.

The steps below summarize the geocoding effort for the infoUSA data. A variety of methods were used to obtain the best possible matches, starting at the parcel level and working up to street network level address ranges.

1. Split out Burlington employment records from infoUSA database - includes zipcodes 05401, 05405 (UVM), and 05406 (PO Boxes).
2. Match remaining infoUSA data (w/o Burlington data) with available e911 esites data. Esites data was processed using the address parser script to allow for better geocoding within the ArcView geocoding engine. Zipcodes were attached to the esites data to allow use of the “single house with zone” geocoding engine [zipcodes were from dynamap statewide zip coverage]. Geocoding was accomplished using the GMI geocoding magician extension for arcview. Offset was 0, and point dispersion was 0.01 miles (this means if a point fell on top of a point it was dispersed 0.01 miles to allow for easier viewing).

5461 records; 2279 matched to esites point data (41.7%). 3182 unmatched

3. Try to match unmatched infoUSA records with municipal parcel information. First attempted to do South Burlington. Cleaned up and made parcel points geocodable. Used [P_loca_a] database field for address information. 115 additional matches achieved.

4. Rematch in areas of the county without e911 esites data available – Huntington, Milton, Westford, and Winooski.

Rematched with Westford parcels. Used E911 database fields for address information [z11 fields]. 26 additional matches achieved (now 3041 unmatched records).

Rematched with Milton parcels. Used [P_loca_a] database field for address information. 169 additional matches achieved (now 2872 unmatched records).

Rematched with Winooski Dynamap roads. Winooski parcel coverage did not have sufficient physical address information to make parcel matching practical. 260 additional matches achieved (now 2612 unmatched records).

Rematched with Huntington Parcel information. Difficult due to lack of physical address information in parcel coverage. 5 additional matches achieved (now 2607 unmatched records).

5. Reviewed unmatched records:

Zipcode

Zipcode	Municipality	Number Unmatched Records
05402	Burlington (PO Box)	4
05403	South Burlington	478
05404	Winooski	14
05407	South Burlington (PO Box)	5
05439	Colchester	2
05445	Charlotte	49
05446	Colchester	499
05451	Essex	19
05452	Essex / Essex Jct	363
05453	Essex Jct (PO Box)	7
05461	Hinesburg	83
05462	Huntington	44
05465	Jericho	104
05466	Jonesville (richmond)	7
05468	Milton	148
05477	Richmond	114
05481	St Albans	1
05482	Shelburne	148
05489	Underhill	31
05490	Underhill Center	13
05494	Westford	21
05495	Williston	453

Corrected unmatched addresses for employers larger than 100

Comparison of infoUSA data to other commercial providers (downloaded 12/8/03):

[http://list.infousa.com/cgi-](http://list.infousa.com/cgi-bin/abicgi/abicgi.pl?bas_session=S23512138200801&bas_elements=4&bas_vendor=190000&bas_type=LC&bas_page=999&bas_action=CompetitiveAudit)

[bin/abicgi/abicgi.pl?bas_session=S23512138200801&bas_elements=4&bas_vendor=190000&bas_type=LC&bas_page=999&bas_action=CompetitiveAudit](http://list.infousa.com/cgi-bin/abicgi/abicgi.pl?bas_session=S23512138200801&bas_elements=4&bas_vendor=190000&bas_type=LC&bas_page=999&bas_action=CompetitiveAudit)

BUSINESS DATABASE COMPETITIVE AUDIT August, 2001 infoUSA • Dun & Bradstreet • Acxiom • Experian

EXECUTIVE SUMMARY

infoUSA contracted with Bass & Associates to conduct an unbiased audit to assess the accuracy of the infoUSA business database and industry competitors. The statistically valid 2001 business competitive audit compared data from infoUSA, Dun & Bradstreet (D&B), Acxiom and Experian in six randomly selected ZIP Codes from various U.S. geographies. The competitive audit consisted of comparing each of the competitor's data sets to infoUSA to determine matching records and unique records. An attempt to verify the information for each business was made by phone to gauge the accuracy for each respective information provider.

The 2001 business competitive audit can be summarized in four points:

1. infoUSA is clearly the industry leader in providing accurate business information compared to D&B, Acxiom and Experian for every category: company name, contact name, address, employment range, phone number and primary SIC code.
2. The file coverage for each participant cannot be assessed by business count alone. Other factors including overall quality, duplicates, out of business records and records without phone numbers has to be considered. Nationally, the file size between infoUSA and D&B is comparable.
3. infoUSA possesses a significant advantage where contact by phone is important (telemarketing, etc). infoUSA experienced a 62% contact rate by phone compared to 55.6% for D&B, 43.2% for Acxiom and 44.3% for Experian.
4. The quality of the D&B, Acxiom and Experian unique records compared to infoUSA is very poor considering the contact rate by phone and out of business rate.

The following summarizes the results of the 2001 audit.

Description	infoUSA	D&B	Acxiom	Experian
Number of Businesses	1,929	2,168	2,684	2,606
Company Name Accuracy	96.4%	96.1%	93.3%	91.0%
Contact Name Accuracy	86.1%	82.2%	80.4%	73.7%
Address Accuracy	95.0%	92.2%	90.8%	92.8%
Employment Accuracy	77.9%	73.8%	73.1%	65.2%
Phone Number Accuracy	98.7%	97.6%	96.4%	96.1%
Phone Numbers Available	100.0%	100.0%	89.7%	84.7%
Primary SIC Accuracy	91.9%	87.9%	81.9%	79.1%
Out of Business Rate	5.7%	9.5%	17.9%	14.2%
Duplicates	2.1%	4.3%	2.3%	3.7%
Contact Rate	62.0%	55.6%	43.2%	44.3%

UNIQUE INFORMATION

The data sets from D&B, Acxiom and Experian were each matched to infoUSA data. A unique record is one that exists in a competitor's file but not the infoUSA database and vice versa. For example, 38.2% of D&B's records were not found in the infoUSA database; conversely, 33.3% of infoUSA's records were not found in the D&B database. The table below includes information related to the quality of information unique to D&B, Acxiom and Experian compared to infoUSA.

Description	D&B	infoUSA	Acxiom	infoUSA	Experian	infoUSA
Unique Records	38.2%	33.3%	55.4%	39.0%	44.0%	27.3%
Contact Rate	38.4%	52.0%	25.3%	54.7%	20.6%	55.9%
Non-Contact Rate	61.6%	48.0%	74.7%	45.3%	79.4%	44.1%
Out of Business Rate	18.8%	10.4%	28.3%	8.0%	25.6%	8.6%