STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE EUROPEAN COMMUNITIES

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

Joint ECE/Eurostat Work Session on Statistical Data Confidentiality
(Skopje, The former Yugoslav Republic of Macedonia, 14-16 March 2001)

Working Paper No. 28/Add.2
English only

ANNEX 2

Quality of Information

Geographic coverage of D&B database

Dun & Bradstreet's Worldbase database is a global marketing database that contains over 53 million business records, including linkage, in 200 countries. Our U.S. marketing database includes 11,300,000 business establishments in the 50 United States plus the territories of Puerto Rico and the Virgin Islands.

Frequency of master database update

Under D&B's Full File Maintenance Strategy, all businesses in the U.S. marketing database are investigated either via a site visit or a telephone investigation at least once a year with many records being touched much more frequently -- as often as several times a month -- through various triggered maintenance programs. The average age of a record in D&B's U.S. business database is 7.5 months.

Procedure for database update

To maintain the dynamic nature of marketing information, D&B utilizes a fully integrated approach towards data maintenance:

D&B's Full File Maintenance Strategy is a planned and integrated full file data maintenance process where all businesses in the U.S. marketing database are investigated either via a site visit or a telephone investigation at least once a year. D&B conducts these scheduled updates. Professionals located in our 50+ field offices around the country or by multiple touch calling from one of our four Telecenter locations.

The D&B Telecenters leverage the latest in predictive dialing technology that links the D&B investigative associates into a fully self-contained scripted calling environment. The Telecenter calling environment contains important validations and cross-checks designed to validate the information the Communicator is gathering while they are still on the phone with the information customer. These efficiencies provide D&B with a capacity to call over 1 million records per month in our Telecenters.

Any updates or additional information gathered during our conversations is automatically fed into daily database update systems. Before any information is updated, those records are subjected to over 2,000 edits and validations including duplicate screening. These validations are in place to ensure our information customers receive only the highest quality information possible.

In addition to the Full File Maintenance Strategy, maintenance investigations can be triggered by many other internal and external information sources such as:

- Private Third Party Data Files
- Public Record Data
- Customer Files
- Newsworthy Services
- Internal Telephone / Mail Activities
- National Change of Address
- Address Standardization Processes

Additional information on these and other triggered programs is as follows:

⇨ Newsworthy Services Group - The group is responsible for monitoring and collecting information from news sources as well as conducting the investigations of these changes before entering them into the database. Changes can consist of business closings, business name change, address change, mergers/acquisitions, etc.

⇨ National Change of Address - Based on information supplied by postal service files, a D&B record will either be excluded from the marketing file after it has been telephone verified that the company is out of business, or an address change will be made. Unlike many other companies, D&B also takes the NCOA "soft-matches" or records the Postal Service is unsure of and telephone investigates to determine if changes have occurred.

⇨ Address Standardization Updates - Address standardization software is used on a daily input as well as during monthly full database sweeps to ensure our addresses are standardized and to apply bar-code and Zip+4 information.

⇨ NewsHound - This is a program where D&B associates report new businesses or changes to existing businesses via fax or phone to a central update location. These changes are then investigated before they are entered into the database.

⇨ Top 1,000 Review - The top 1,000 companies in the information base by sales and employees are manually and machine reviewed for inconsistencies. Updates are applied as needed.

⇨ Linkage Updates - All "family trees" are updated on a 12 month cycle using a "top-down" confirmation of the headquarters locations. Additionally, all branch locations are investigated in the same 12 month cycle using a "bottom-up" confirmation of the marketing information specific to that branch location and the linkage to its headquarters is checked.

⇨ On-Demand Programs - In addition to all of the scheduled maintenance programs, D&B recognizes some changes which occur to businesses that need to be updated in the information base right away. For example, business moves, mergers, bankruptcies, or CEO changes which D&B learns of will trigger an immediate investigation by a D&B analyst.

Formal process

Prior to adding records to the database, investigations are conducted at either one of our 50+ field offices or at one of our four Telecenters where our Communicators work using state of the art technology that includes a scripted calling environment. These investigations serve to confirm information as well as gather additional information such as contact names, number of employees, year started, linkage data, tradestyles, additional address information, home based status, just to name a few. Investigations are conducted during normal business hours as well as evening and Saturday calling.

Each record is then subjected to over 2,000 edits and validations including screening for duplicate information against the entire database prior to inclusion in the marketing database. A team of associates is then responsible for resolving all error conditions via automated process flows and/or reinvestigation. The new business listings are then entered into our database daily using fully automated systems.

Source of data elements

Dun & Bradstreet leverages a variety of sources for new record identification, including:

⇨ Directly from the business principal during investigation.

⇨ Response to inquiries from D&B credit customers.

⇨ Customer Files – D&B partners with many of its' customers to identify new businesses and add them to the marketing file after investigating and enhancing the information.

⇨ Personalized Investigations – These new businesses enter our database as a result of customer requests for specific investigations.

⇨ Public Record Sources (Local, State and Federal) - Including New Business Registrations, Corporate Charter Details and Public Bulk Source Files.

⇨ D&B Bounty Program – This program utilizes our associates in over 50+ field offices located throughout the country by motivating them to proactively identify new business start-ups in their area.

⇨ Private Third Party sources – Many of these files are acquired via niche specific compilers of information.

⇨ Business Directories

Please see Attachment A for source information on the specific data elements D&B is recommending for your company.

Quality of data

Marketers require the most-up-to-date information in targeting their markets and reaching their most likely prospects, in order for their marketing campaign to be cost-effective. D&B has numerous quality checks in place to help ensure this. Some of the programs include:

⇨ Quality Review Telecenter Output files

⇨ A system of over 2,000 Edits and Validations, which occur between our master database (AOS), and the marketing database at the data element level. These edits and validations are cross-relational: For example, they compare the listed city, state and zip code to ensure valid address information and they compare sales, number of employees and SIC to identify businesses with abnormally high figures for a given industry.

⇨ Address Standardization is performed by using a software program that we receive from Group One called Code 1 Plus. This software program is run on a monthly basis to validate and correct 5-digit zip codes, as well as assign Zip+4 extensions. Address validation is also performed along with city correction and delivery point assignments for bar-coding.

⇨ A variety of data cosmetic procedures are implemented to synchronize upper/lower casing, invert business names into standard format (appear as John Smith Industries in the marketing database), and CEO name parsing for genderization and letter personalization.

⇨ Monthly audits are performed since most of the edits and validations occur on an individual record-by-record basis, a monthly audit and standard report are created. Frequency counts, distribution analyses, and profiles are produced and compared to the previous month's results. Unusual occurrences or abnormally high or low frequencies are identified for investigation and correction.

⇨ Accuracy Measurements of the marketing database are performed on a quarterly basis through audits. These are performed by using statistically valid sampling techniques to extract marketing database records, which are then mailed and telephoned, to achieve accuracy levels of plus or minus 1%. These results are then further analyzed to identify segments of the file that may need additional attention.

⇨ Customer Feedback, we are eager to receive feedback from our customers on ways to improve the quality and content of the marketing database. This feedback is often incorporated into change activity targeting the improvement of procedures, processes and overall data integrity.

⇨ Area Code changes in the marketing database begin with a monthly receipt of data from BellCore, the company that manages the North American phone numbering system. Approximately 30 days before the split is to occur, all of the exchanges moving to the new area code are pulled from the BellCore file. These exchanges are put on a tape, which is processed at the National Business Information Center (NBIC), effectively changing all of the records in our database (AOS). This is performed on the same weekend as the split actually occurs. The marketing database is then fed these changes from AOS and they are included in the subsequent cut of the marketing database. The timing of these splits is the determining factor for when the new area codes will appear in the marketing database. For most splits, the next update of the marketing database will contain the new updated area codes. In addition, at the end of the permissive dialing period for a particular split, another sweep is performed by NBIC to change any records that do not reflect the new area code and validation tables are updated in the data entry system to check the new area code and exchange combination.

⇨ Each month the United States Post Office makes changes to existing ZIP Codes, creates new ZIP Codes or discontinues existing ZIP Codes. The marketing database is swept each month to update the ZIP Codes on our records with the USPS changes.

⇨ D&B utilizes a product from Group One Software called Code One Plus to standardize street addresses and update ZIP Code changes within the database. Group One receives the postal changes

from the USPS each month. Group One then formats the new information to be compatible with the Code One Plus software package and sends us monthly updates. Before the marketing database is cut each month, the software is used to update the ZIP Codes.

Percentage of frequency each data element

Please refer to your custom data layout for frequency information on the data elements D&B is recommending for your company.

Percentage of data elements that are "actual" vs. "modeled"

Please refer to your custom data layout for source information on the data elements D&B is recommending for your company.

Industry Classification

Dun & Bradstreet assigns SICs at the eight-digit level and can include up to six base SICs, prioritized based on the revenue contribution of each business activity, with up to four four-digit extensions for each, for a maximum total of twenty-four separate line of business assignments per record.

To help address our customers' need for a more definitive and dynamic system, Dun & Bradstreet developed an enhanced SIC system in 1986, resulting in the SIC 2+2 coding scheme. By appending two new digital (hierarchical) categories to the existing structure, SIC 2+2 builds upon the government and expands the number of classifications nearly eighteen-fold.

D&B's SIC 2+2 is an eight-digit number that consists of two four-digit parts:

(1) A government four-digit SIC, known as Base 4, which describes business activity at an industry level, and

(2) A D&B four-digit SIC, known as the Extension 4, which consists of two two-digit parts (2+2). The first two-digit component is a sub-industry code which represents products or services which share common characteristics. The second two-digit component is a line of business code which represents specific products manufactured or specific activities performed by a business.

Note: The Federal Government began replacing the 60-year old Standard Industrial Classification (SIC) system with the North American Industry Classification System (NAICS), a new and improved system, which offers substantial structure improvements and identifies emerging industries. However, at the eight-digit level, D&B's SIC 2+2 still allows for a much more detailed description that NAICS and will continue to better identify traditional industries.

NAICS will be the new uniform standard industry classification system created by three NAFTA countries for implementation in 1997. It will replace classification systems now being used in the United States, Canada and Mexico. NAICS is a six-digit coding system which may only go to five digits in some instances.

The SIC system was based on basic concepts developed in the 1930's when our economy was primarily manufacturing oriented. The last major change was in 1987, which enabled Dun & Bradstreet to create its SIC 2+2 system. In 1992, plans were made to create a new system to increase service orientation, recognize new and emerging industries and react to the public's dissatisfaction of the SIC system. The new NAICS system is more process and employment based than the old system which leads to a more consistent method for classifying businesses.

The old SIC system is not going away, but the Federal Government will no longer support and maintain the system.

Dun & Bradstreet has received a copy of the new system and is building a crosswalk mapping SIC 2+2 codes to NAICS Codes. This crosswalk will be used in custom orders for those customers that wish to append NAICS codes to their marketing orders. D&B associates will continue to assign SIC 2+2 codes. Therefore, customers will be able to continue to use and select records based on an SIC code.

Detail of corporate family linkage

The D&B D-U-N-S® Number, which will be appended to all matched customer records, will enable your company to effectively aggregate all customer sites to an enterprise level.

The Data Universal Numbering System (D-U-N-S) is a non-indicative, special nine-digit number that D&B assigns to each unique business location in the D&B information base, giving you a more accurate, consistent way to identify businesses in your files and transaction systems. The first eight digits are sequentially machine-generated; the ninth digit is a "check digit" to catch input errors. The D&B D-U-N-S® Number is assigned and maintained only by Dun & Bradstreet.

The D&B D-U-N-S® Number is the worldwide standard for business identification and is endorsed by the United Nations, International Standards Organization, European Commission, U.S. Federal Government and over 50 industries for EDI use, including requirements from the Federal Energy Regulatory Commission (FERC) in FERC Orders 889 and 563-A.

Dun & Bradstreet's company family linkage, which is made possible by the D-U-N-S® Number and will also be appended to all matched customer records, will enable your company to effectively roll up locations to customers. Unlike our competitors, Dun & Bradstreet's linkage can support unlimited family depth (layers of ownership and reporting) and breadth (total number of family members), both domestically and globally.

Dun & Bradstreet's domestic linkage file is derived from the U.S. Marketing Database and the D&B Canadian file and is called the Company Affiliations File (CAF). The CAF file links all members of a company "family tree" including any establishment that is directly or indirectly linked through ownership of more than 50% (parent/subsidiary relationship) and headquarter/ branch relationships. This file is updated on a monthly basis by a dedicated staff of family linkage experts. D&B linkage does not include organizational linking of neither executives nor linkage of affiliate companies with less than 51% ownership.

Linkage, in general terms, is the relationship between different companies within a company family. Headquarters, branches, single locations, parents and subsidiaries are all viewed in terms of their relationship to other family members of the same family. The basic types of records that are included in linkage are:

Ultimate

An ultimate is the highest ranking company in a company family structure. It encompasses all directly related branches, subsidiaries and parents of a specific business. The ultimate is the key to isolating and organizing family trees. In general terms, the ultimate company can be identified when the Case ID D&B D-U-N-S® Number is the same as the Ultimate D&B D-U-N-S® Number and the hierarchy code equals "01".

Single Location

A single location is a business entity which has no other unique D&B D-U-N-S® Number or location reporting to it. (Do not confuse with 'stand alone' business, which are not linked in a family).

Headquarter

A headquarter is a business entity which has at least one branch reporting to it. When a business has multiple branch locations, the designated headquarter location is determined by the controlling management of the business.

Branch

A branch is a secondary location for which its headquarters has legal responsibility. Typically, a branch is at a separate location. However, a branch can be located at the same address with its headquarters or sister branch, provided that they have unique, separate and distinct operations. Branches often have secondary names or tradestyles, but always carry the same primary name as their headquarters. A branch record must be linked to a Headquarters D&B D-U-N-S® Number.

Division

A division, like a branch, is a secondary location of a business. However, a division carries out specific business operations under a division name. Divisions look similar to branches since they also carry the branch status code.

Parent

A parent is a business entity with controlling interest in another company through ownership of a majority (more than 50%) of its stock. The parent-to-subsidiary relationship is similar to a headquarter-to-branch relationship with the exception of its legal structure.

Subsidiary

A subsidiary is a corporation with more than 50% of it stock owned by another business. A subsidiary may be located at the same address as its parent.

Each record also carries a set of linkage elements which help to identify the type of record it is, as well as its relationship to other records in the family tree. These linkage elements are the Status Code, Subsidiary Code, Hierarchy Code, and Dias Code.

The *Status Code* is a one digit field that identifies a record as:

- ⇨ Single Location 0
- ⇨ Headquarters 1
- ⇨ Branch 2

The *Subsidiary Code* is a one digit field that identifies the record as:

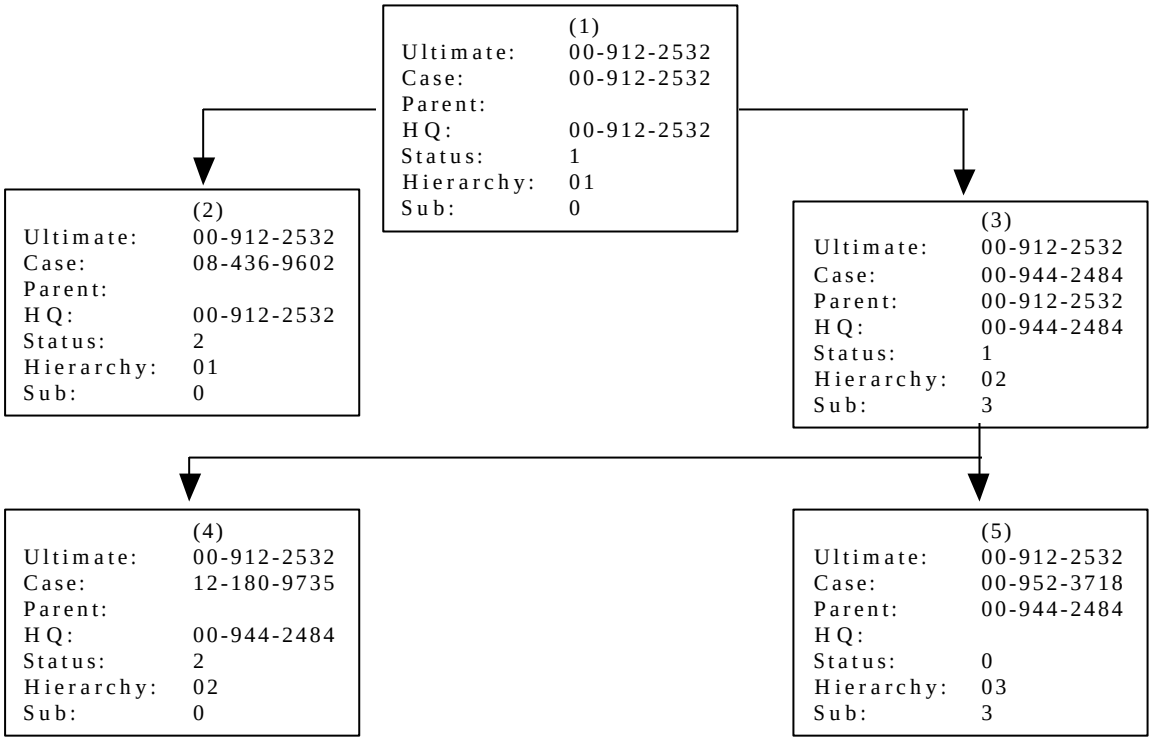
- ⇨ Subsidiary 3
- ⇨ Non-Subsidiary 0

The *Hierarchy Code* is a two digit field that determines the records relative position in a family tree by indicating its relationship to the other records.

- ⇨ Valid Code Values are 01-12

The *Dias Code* is a single nine-digit field that sorts a corporate family into family tree sequence. In general terms, all branches will be listed directly below their headquarters while the subsidiaries will be listed directly under their parents. In a situation where a parent/headquarters has both branches and subsidiaries reporting to it, the branches will be listed first, followed by the subsidiaries. Branches are sorted by state, while subsidiaries are sorted by company name. Note: The Dias code changes each time the linkage file is updated.

Family Tree Example



Quality of the linkages

All business establishments in a family tree are updated on a 12 month cycle. During that update, business information specific to that location is gathered and the linkage to the next highest level is confirmed.

To insure the quality of this critical information, Dun & Bradstreet has also implemented a Central Linkage Review Process. Each year, a group of 60 highly trained Dun & Bradstreet professionals dedicated to the linkage updating process complete a "top down" review of the largest 4,000 company family trees. During the linkage review, D&B interviews the controller, legal counsel or public relations executive of the ultimate parent company to verify the relationship of all subordinate business establishments.

Dun & Bradstreet is committed to providing our customers with the highest quality linkage information possible and currently spends approximately 22% of the total data collection and maintenance budget on company family linkage.

Customer Service and Support

Dun & Bradstreet is committed to supporting your company in maximizing the value derived from D&B's information and services. Dun & Bradstreet's team of local, regional, headquarters and industry experts are available to support your company in thoroughly understanding the data content, data quality and business applications for the business database and related services.

In addition, Dun & Bradstreet has a staff of experts available for consulting, on a contract basis, in the following areas:

⇒ Database Marketing – Data Integration, Systems Design and Implementation

⇒ Analytical Services – Customer/Market Analysis, Demand Estimation, Model Development