

## THE DEVELOPMENT OF THE MASTER ESTABLISHMENT LIST

David Hirschberg, Small Business Administration

As part of its data base developmental effort, the Office of Advocacy, Small Business Administration (SBA), has developed a Master Establishment List (MEL) with over 8.1 million businesses. In creating the list, two commercially available lists were merged. The first, the Dun's Market Identifier file, contained over 4.6 million records; the second, the Market Data Retrieval file--a "yellow-page" listing--contained over 7 million records.

The MEL provides direct statistics on the number and geographic distribution of America's small businesses. It also facilitates communication with the small business sector and is a vital tool for conducting surveys and mailings to selected industrial sectors regarding governmental policy.

This paper describes the development of the Master Establishment List. First, some background is provided on existing small business files. Then the MEL is discussed, some of its uses are described and some on-going validation efforts are mentioned. The paper concludes by raising some of the policy implications of concern to SBA.

### BACKGROUND

Although major progress has been made, the small business sector remains poorly documented in the Federal statistical system. Most existing Federal statistical data and administrative record sources are not adequate for assessing the impact on small business in a variety of policy analysis and decision-making areas. It is interesting to note that of the 124 pages of statistical tables appearing in the Economic Report of the President, 1985, only one is relevant to small business activity, "Business Formation and Business Failures, 1940-84." [1] (The source of this business formation and business failure data is Dun and Bradstreet.) Two other sources of information on business formation are the Bureau of Economic Analysis and the Internal Revenue Service. However, there are obvious problems in using their data

as well. For example, the Index of Net Business Formation, published by the Bureau of Economic Analysis, is 114.8 for 1983 (with 1967 = 100). This growth level is sharply at variance with the number of business tax returns reported by IRS, as shown below. Furthermore, the number of enterprises has increased from 3.3 million in 1976 to 4.4 million in 1982.

The Small Business Administration, Office of Advocacy's Small Business Data Base was designed to provide more reliable information on the scope and contribution of the small business sector. This data base is drawn from commercially available data and places little additional paperwork burden on the business community. It permits the maintenance of confidentiality and provides policy-relevant data.

The first project, which is now complete, was the development of the United States Establishment and Enterprise Microdata (USEEM) files for 1976, 1978, 1980 and 1982. These files are based on Dun and Bradstreet's Market Identifier (DMI) files, which are collected for credit and insurance purposes. They have been edited, cleaned and reformatted, and are the basic centerpiece of the Small Business Data Base.

These four files contain information on business organizations that reported business activity in any one year. Each record which identifies an establishment has the following information: (1) Dun's number--this is a number assigned by Dun and Bradstreet that uniquely identifies each establishment and can be used to merge with prior-year files; (2) geographic location -- city, county, SMSA, state and zip code; (3) year business started; (4) number of employees; (5) annual sales volume; (6) Standard Industrial Classification (SIC) code; (7) parent and headquarter's city and state; (8) Dun's number of parent and ultimate parent; (9) subsidiary indicator; (10) status indicator -- single location, headquarters, establishment or branch; and, (11) manufacturing indicator -- indicates if manufacturing takes place at the location.

Table 1. IRS Business Tax Returns by Legal Form of Organization  
(in millions)

Year	Total	Proprietorships	Partnerships	Corporations
1967	8.5	6.1	.9	1.5
1976	11.3	8.1	1.1	2.1
1982	14.6	10.2	1.5	2.9

Source: Statistics of Income Division, IRS.

The USEEM files now contain data for the estimated 8 million business establishments which existed during the period 1976-82. For each year, annual files include approximately 5 million records. These records provide estimated employment and industry classification for establishments and firms, the start date (age), organizational status and geographic data for each firm.

These USEEM files have been linked into a longitudinal sample file, the United States Establishment Longitudinal Microdata File (USELM), enabling researchers to follow the same establishments over time. This is a primary and necessary requirement to address policy-relevant research issues. The 1984 files are currently being developed; they will later be merged with the USELM 1976-82 files.

The second project involves working with Dun and Bradstreet's raw financial statement file (FINSTAT). The FINSTAT file contains about 150,000 financial statements for 1975, but for the past few years the number has increased to almost 500,000 per year. To preserve the confidentiality of cooperating companies, all identifying information has been removed by Dun and Bradstreet. Although the file includes the major U.S. corporations, approximately 95 percent of the firms have fewer than 100 employees and 74 percent have fewer than 20 employees. By comparing these data with other sources, we are beginning to resolve the question of how well these data represent the small business community.

Finally, a major effort is underway to have data available on small business from the various statistical and administrative agencies of the Federal Government. Together with the Internal Revenue Service (IRS), for example, the Small Business Administration is supporting an effort to link IRS' business Statistics of Income files for partnerships, proprietorships and corporations with that agency's tax reports of employment and payrolls. This overcomes a significant shortcoming in the IRS files. As rich as they are for analytical purposes, there is no employment reported on business tax returns. Other projects include organizing the IRS Corporate Source Book [2] into machine-readable form and examining disclosure and confidentiality issues, particularly as they relate to business data from IRS and Census sources, so as to develop disclosure strategies for the release of microdata (data on individual firms).

#### THE MASTER ESTABLISHMENT LIST (MEL)

A universe list of firms and establishments is the core element of a statistical program. The Bureau of the Census uses the annual IRS business tax returns, combined with employer withholding/social security reports and multi-establishment company surveys, to develop their list of businesses with employees--the Standard Statistical Establishment List (SSEL). Multi-establishment companies of the Company Organization Survey enable the SSEL data to provide linkage between establishments and their parent firms. The total number of establishments in the SSEL in 1977 was approximately 4.3 million,

compared with the 15.6 million business tax returns. Most of this difference is made up of firms without employees.

The Bureau of Labor Statistics (BLS) also prepares lists of establishments or, more correctly, tax units. Administrative records from each of the State unemployment insurance systems are compiled annually. Linkages between the establishments and their enterprises are not available. Other agencies have developed lists to meet their needs as well. An example is the Post Office/Survey Research Center Sample of Nonhousehold Mailers.

Unfortunately, Advocacy cannot use the Census, IRS, or BLS lists as the basis of its sampling frame. By law, the information in these sources cannot be disclosed. Therefore, Advocacy undertook to develop a Master Establishment List based on merging two publicly available private sources: (1) the Dun and Bradstreet's Market Identifier (DMI) file and (2) a "yellow-page" listing from Market Data Retrieval, Inc. (MDR) for the year 1981. The MDR file is compiled from 9 million entries, including duplicates, in the nation's telephone directory yellow pages. The MDR covers many of the establishments in the DMI file and also many small establishments and persons who do not have credit ratings.

Merging the DMI and MDR files involved a considerable effort, given the enormous size of these files and the absence of unique identifiers. [3] About 3.5 million unduplicated records in the MDR file were identified as not having a matching record in the DMI file. The resulting MEL file contains a total of 8.1 million firms and establishments for 1981. [4]

The coverage of the MEL is important. It is useful to compare with comparable tabulations of employment from the Census Bureau's County Business Patterns (CBP). Table 2 does this for the DMI components of the Master Establishment List.

The first two columns of Table 2 list the number of establishments identified in the DMI and CBP. As mentioned previously, the DMI file covers all establishments with Dun and Bradstreet credit ratings. This includes a small number of establishments with no employees, as well as an undetermined number of small establishments with employees. In contrast, the CBP includes only establishments with employees. Given these coverage differences, it is noteworthy that there is a basic similarity in the total number of establishments.

Several reasons exist for the differences by industry, but they are difficult to quantify. Discrepancies may result from differences in industrial classification between the DMI and the CBP. The extent to which the DMI file includes firms with no employees, as well as establishments which are no longer in business, is not known.

Given these classification and coverage problems, the employment estimates are remarkably similar at the major industry division level, as shown in Table 3. Total employment in the DMI file is 6 percent less than that of BLS and 2 percent more than that of CBP. For mining, contract construction, manufacturing, and services, the DMI reports slightly more employment

Table 2. Establishment Counts by Major Industry Division: Dun's Market Identifier (DMI) and County Business Patterns (CBP), 1981

(Establishments in Thousands)

Industry	DMI	CBP	Ratio DMI/CBP
All Industries, Total	4,635	4,587	1.01
Agriculture, Forestry & Fishery	120	804	.15
Mining	42	359	.12
Construction	612	626	.98
Manufacturing	441	336	1.31
Transportation, Communications & Public Utilities	182	162	1.12
Wholesale Trade & Retail Trade	1,846	1,887	.98
Finance, Insurance & Real Estate	372	387	.96
Services	1,019	1,445	.71

Note: Components may not add to total due to rounding.

Source: Tabulations from the DMI and County Business Patterns, U.S. Bureau of the Census (selected years).

Table 3. Employment by Major Industry Division: Dun's Market Identifier (DMI), Bureau of Labor Statistics (BLS) and County Business Patterns (CBP), 1981

(Employment in Millions)

Industry	DMI	BLS	CBP	Ratio		
				CBP/DMI	BLS/DMI	BLS/CBP
All Industries, Total	74.7	75.1	74.8	1.001	1.005	1.004
Agriculture, Forestry & Fishery	.8	NA	.3	.38	NA	NA
Mining	1.3	1.1	1.1	.85	.85	1.00
Construction	5.9	4.2	4.3	.73	.71	.98
Manufacturing	21.2	20.2	20.4	.96	.95	.99
Transportation, Communications, & Public Utilities	4.1	5.2	4.6	1.12	1.27	1.13
Wholesale Trade & Retail Trade	16.7	21.6	20.3	1.22	1.29	1.06
Finance, Insurance & Real Estate	4.6	5.2	5.4	1.17	1.15	.98
Services	19.0	18.6	17.9	.94	.98	1.04

Note: Components may not add to total due to rounding.

Source: Preliminary Report on the Development of the Master Establishment List, 1982, Social and Scientific Systems, Inc.

Table 4. Dun's Market Identifier (DMI) and Market Data Retrieval (MDR) Files as Components of the Master Establishment List, 1981

Number of Establishments in Thousands

Industry	DMI	MDR	MEL	Ratio MDR/DMI
All Industries, Total	4,635	3,488	8,123	.75
Agriculture, Forestry & Fishery	120	49	169	.40
Mining	42	10	52	.25
Construction	612	215	828	.35
Manufacturing	442	82	524	.19
Transportation, Communications & Public Utilities	182	84	267	.46
Wholesale Trade & Retail Trade	1,846	1,054	2,900	.57
Finance, Insurance & Real Estate	372	407	779	1.09
Services	1,019	1,577	2,595	1.54

Note: Components may not add to total due to rounding.

Source: Preliminary Report on the Development of the Master Establishment List, 1982, Social and Scientific Systems, Inc.

than the CBP or BLS files. However, there is significant undercoverage for wholesale and retail trade; transportation, communications and public utilities; and finance, insurance and real estate.

Unfortunately, employment is not available from the MDR file, but the number of establishments added to the DMI file is shown in Table 4. It was apparent from the detailed industry tabulations that the added MDR firms were mostly professionals, such as doctors and lawyers, as well as taxi operators, truckers, insurance agents, and real estate brokers -- businesses that generally do not use credit. These sectors are basic to small business activity and it is important that they be included in lists of small businesses.

In contrast to the 15 million tax returns filed with IRS, the Master Establishment List contains 8.1 million firms and establishments. It does not follow that there is a deficiency in the MEL. Inspection of the sales distribution reported in IRS' proprietorship files suggests that they include persons with other occupations and do not truly reflect full-time business activity. Of the 12.7 million proprietorship reports in 1980, almost half have business receipts below \$5,000.

The analysis of the DMI file and the business units added by the MDR file indicate that, for most purposes for which the file will be used, the MEL is representative of the full-time business population with employees.

#### USES OF THE MEL

The Master Establishment List has been used for a variety of purposes. Users studying specific problems relating to small business have requested that the Small Business Administration make specialized tabulations from the MEL, draw samples based on those tabulations, and provide mailing lists for the sample cases. In some cases the requests have asked for firms by industry and size for a specific State or designated SMSAs or even particular counties. Although some users have been concerned with the broad spectrum of business units, other users' interests have been highly specialized.

An example of the use of the MEL to create a specialized data base was its use in analyzing the proposed legislation on enterprise zones. Because the establishments in the MEL have addresses, it is possible to examine the existing location of business activity in central cities and non-central cities in relation to the proposed enterprise zones. Some measure of the magnitude of potential costs and benefits of the legislation can be obtained by analyzing projected changes in business activity and employment.

In another application, using a three percent sample of the MEL's businesses, an Ownership Characteristics Survey was initiated in January of 1984. It asked respondents for the legal form of ownership as well as for the sex, race and veterans status of the business owner.

Summary results are available in the "Report of the President on the State of Small Business, 1985." [5]

#### VALIDATION EFFORTS

The exact matching of the 4.6 million DMI records and the 9 million MDR records to produce 8.1 million Master Establishment List records was considerably more successful than might have been expected, and the resulting MEL file has had wide use. As the tabulations of MEL show, the DMI data were augmented in precisely those areas where it was known that coverage was incomplete (i.e., services and trade). Although there are undoubtedly additional small businesses that are without Dun's credit ratings and are not listed in the yellow pages, it is not clear that further efforts to extend the MEL would be worthwhile.

Validation studies have been carried out analyzing the MEL's coverage, consistency, and completeness. One such study involved matching the establishments in the area samples of the University of Michigan's Survey Research Center with the establishments listed in source areas in the Master Establishment List. Another study is comparing State unemployment insurance data with DMI files.

The former study revealed important differences in the MEL list and the list compiled by Michigan. However, recent research has indicated that these lists are subject to obsolescence. Turnover is about one percent a month; therefore, if lists compiled for different time periods are compared, a large number of nonmatches should be expected. This and other experience has shown that a large proportion of nonmatches occurs when business lists are matched using different sources of information. [6]

In the latter study, unemployment insurance microdata files and DMI files were matched for a recent time period for Texas and Pennsylvania. When the comparisons are completed, they will yield information of considerable value in evaluating the DMI file. It can be noted that only about 40 percent of the firms in the files were matched.

#### FEDERAL POLICY IMPLICATIONS

Using the January 1985 DMI and MDR files, an updated MEL is being created. We are asking support from the various statistical agencies to provide resources to continue this effort, to improve its quality and help make it generally available to the statistical community.

There is a clear need throughout the Federal establishment for a consistent and reliable

business universe frame for a variety of research and sampling purposes. Each Federal agency now operates its own system, virtually oblivious to the activities and requirements of others. Employment differences between systems are explained as due to classification, reporting and coverage procedures. In this time of considerable budgetary restraint, cooperation in the development of databases such as the MEL is absolutely necessary.

#### ACKNOWLEDGMENTS

The development of FINDIT and the actual computer matching were carried out by Social and Scientific Systems, Inc., a Maryland based 8(a) firm. The contributions of Dennis Abels, Wendy Alvey, Sam George, and Pal Khera are gratefully acknowledged.

#### NOTES AND REFERENCES

- [1] Council of Economic Advisors. (1985) Economic Report of the President, 1985, table B-91, p. 337.
- [2] The Internal Revenue Service's Statistics of Income Division produces a Corporate Source Book annually, which presents detailed income and balance sheet data classified by industry and size of total assets. For more information, contact the Corporation Returns Analysis Section, Statistics of Income Division (D:R:S:C) at IRS.
- [3] For a detailed description of the methodology and computer algorithm, see "File Matching Utilizing Automated Heuristic Techniques (FINDIT)," by Social and Scientific Systems, Inc., Bethesda, MD, January 1983.
- [4] See "Preliminary Reports on the Development of the Master Establishment List," by Social and Scientific Systems, Inc., Bethesda, MD.
- [5] Small Business Administration. (1985) The State of Small Business: A Report of the President, 1985.
- [6] Converse, Muriel and Heeringa, Steven G. (1984) "An Evaluation of the Accuracy and Current Utility of the 1981 Master Establishment List (MEL)," Institute for Social Research, University of Michigan, Ann Arbor, MI.