

INDUSTRIAL DIRECTORY - STATUS AND DIRECTION

Michael G. Farrell and John M. Sullivan, U.S. Bureau of the Census

I. Overview

The Industrial Directory is a standardized computer file, or register, of all United States business firms and their establishments. It is compiled, maintained and used by the Bureau of the Census on a continuous basis. The Directory contains basic economic information for some five and a half million corporations, partnerships, sole proprietorships and other businesses with employees. Its scope spans all economic activities with the exception of private households. Sufficient information is available in the Directory to provide for the selection of stratified statistical samples, preparation of name and address mailing lists, tabulations of statistical aggregates, use of common Standard Industrial Classification (SIC) codes, use of geographic codes that locate establishments in specific local areas, and other capabilities expected of a universal sampling frame.

For each organizational business unit, the Directory contains the latest information available on: primary name of the firm; secondary name of the firm; mailing address; actual address or physical location; geographic codes; Standard Industrial Classification code; enterprise code; legal form of organization; employer identification number; Directory identification number; total 1st quarter employment, as of the pay period including March 12; total annual payroll; sales and receipts.

Although actual data values are maintained for employment, payroll and receipts, they are represented in the Directory by a set of standard size codes indicating value ranges. This ensures that no disclosure of data belonging to an individual company will occur through use of the Directory. Employment and payroll values are maintained for both the current year and the previous year; sales and receipts data currently are available only in economic census years.

Directory information for another six and a half million non-farm firms without employees is compiled once each five years, at the time of the economic censuses. These nonemployers, typically small businesses operated by sole proprietors or working partners, account for less than 2 percent of business receipts. Because they are relatively less significant than employer businesses, the cost of annually maintaining this segment is not warranted.

The separate Farm Directory also is assembled once in each 5-year period, using results from the Census of Agriculture. Most of the two and a half million farms comprising this list are nonemployers identified during census processing.

The Directory has been developed in answer to specific needs. Under the decentralized statistical system which has developed in the United States, each statistical government agency or unit is responsible for conducting the data collection activities related to its programs. Each agency compiles and maintains lists of firms and establishments to collect economic statistics and classifies these firms and establishments by industry or kind of business. Although each agency presents its data in terms of the Standard

Industrial Classification (SIC) system, the actual code assigned depends on the amount of information available to the agency and the interpretation of that information. The outcome is a series of different lists of establishments for what is supposed to be the same industry. The lack of agreement concerning the identity and industrial classification of the establishments representing the industry in question seriously affects comparability of the data provided by the agencies. Reconciliation and coordination of various statistical series are often difficult and sometimes impossible. Moreover, independently developed lists, with attendant differences in definition and coverage, have resulted in considerable duplication of effort, processing inefficiencies, and increases in respondent reporting burden. Concerns such as these constitute a substantial part of the criticism of government statistical programs.

The concept of a centralized Directory to address these concerns has a long history of endorsement. Eight separate recommendations have called for a central register system to be implemented, starting in 1937 when the Committee on Government Statistics and Information Services (a committee of the American Statistical Association and the Social Science Research Council) recommended that a consolidated mailing list of businesses be maintained. The list as called for would also be used for sample selection and would promote data comparability and uniform coverage.

Subsequently, recommendations in favor of the concept generally or the Census Bureau's Directory specifically include: 1949 - Commission on Organization of the Executive Branch of the Government; 1954 - Report of the Intensive Review Committee to the Secretary of Commerce; 1961 - Report of the Subcommittee on Economic Statistics of the Joint Economic Committee; 1971 - Report of the President's Commission on Statistics; 1977 - Report of the Commission on Federal Paperwork; 1979 - Federal Statistical System Reorganization Project; 1979 - General Accounting Office Report on the Standard Statistical Establishment List.

In October 1968, the Bureau of the Budget, now the Office of Management and Budget (OMB), designated the Bureau of the Census as the focal agency for developing and maintaining a directory of U.S. business enterprises and their establishments. Objectives of the system were at a minimum to:

- Improve significantly the comparability of published data series by providing a single source for the determination and assignment of the Standard Industrial Classification codes of establishments and companies engaged in economic activity;
- Reduce the duplicative data collection costs to government and the duplicative reporting burden on the business community;
- Provide a current file of the business universe organized within an accessible statistical framework so that scientific statistical samples stratified by industry, area, size, or other available economic characteristics can be drawn. It should provide flexibility so that samples

can be selected at the fully consolidated enterprise, unconsolidated legal entity, or individual establishment level;

- Develop as an integral by-product of the Directory system, an annual publication program presenting the statistical information compiled as part of the Directory operations;

- Provide the latest additions or changes to the file at periodic intervals to major statistical agencies.

In 1972, funding for the project started. Developmental work progressed rapidly and included a series of interagency meetings to determine scope and content, and to resolve operational and legal considerations. The Directory became operational shortly thereafter (producing the first County Business Patterns publication on an establishment basis for 1974) and currently is used successfully as the source file for the Bureau's economic censuses and current surveys. The Directory cannot yet be used by other agencies, however, because of the confidentiality provisions of title 13 of the U.S. Code.

II. Background

The building block of the Directory data base is the establishment. An establishment is a single physical location where business is conducted or where services or industrial operations are performed. It is generally considered to be the smallest basic unit for which key figures of economic data, such as employment, payroll, and the value of products or services are available. The use of the establishment concept provides for a more detailed and definitive level of data reporting and publication not possible under other reporting unit concepts which typically combine data for two or more separate establishments. The establishment is useful not only for collecting and presenting data on an industrial and geographic basis, but also provides for the building up to broader organizational levels through aggregation of establishment level detail.

This ability to link together and identify the affiliation of parent companies, subsidiary firms and their establishments is built into the Directory ID numbering system. It includes use of the Internal Revenue Service Employer Identification (EI) number and the Directory file number to identify uniquely three types of records for each multiestablishment firm: a record for the enterprise or parent company; a record for each legal entity or subsidiary company; and a record for each establishment operated by the company. The legal entity level is an organizational unit which, for tax reporting purposes, is assigned an EI number by the Internal Revenue Service. The legal entity or EI may comprise one establishment or many establishments, depending on the company's tax reporting structure, but each legal entity has a separate EI number.

For single establishment companies, the enterprise, legal entity or EI and establishment are synonymous and require one record in the Directory.

This structure and numbering system has evolved to meet several needs. For multiestablishment firms, the ability to identify the affiliation of parent companies, subsidiary legal entities and their establishments provides for: the collection and publication of data for enterprises per se, in addition to their individual establishment level

data; ensuring that there is no disclosure of data for individual companies by linking all affiliated companies under a master number; mailing, collection and correspondence related to report forms that can be directed to the company level or to specific subsidiaries or establishments depending on the needs of the program; a base for merger, acquisition and concentration studies.

Also, and of critical importance, is the ability of the Directory structuring and ID system to link directly with, and draw from, the administrative record systems of the Internal Revenue Service (IRS) and Social Security Administration (SSA). The administrative records of these agencies, which result from the payroll withholding and business related income tax collection processes, essentially define the universe of legal entities to be included in the Directory. They offer an efficient, low cost alternative to the direct mail canvass of millions of firms that would otherwise be necessary to establish the Directory.

Information for all single establishment companies is derived from administrative records, which includes the formation of new businesses, business deaths, 4 quarters of payroll, 1st quarter employment, names, address, EI numbers, legal form of organization, industrial activity codes when not available from earlier censuses, etc.

Administrative record data are the only source of a virtually complete list of companies engaged in all types of economic activity. Lacking this resource, it would be necessary to establish lists independently with no mechanism to ensure complete or even representative coverage. Although incomplete lists are available from private sources and from other government agencies, they lack classification information for specific locations and are almost completely void of information on smaller firms. It is for these millions of smaller firms that experience has shown the response rate of statistical reporting and the quality of the data reported are extremely low. IRS data provide a degree of coverage and quality for this part of the business community that would be impossible to duplicate by other means. Further, even for the large companies that are directly canvassed, the IRS data are an invaluable tool used in the editing and refining of reported data, and used to ensure the completeness of the reports.

These uses of business related administrative record information (such as universe definition, proxy data for single establishment firms, control totals, editing tools and the like) not only reduce respondent burden significantly and increase the efficiency and quality of the Directory, they also lower the cost of Directory operations.

Administrative record data are not without their disadvantages, however. For example, the reliability and level of detail of industrial activity codes available from administrative data are sometimes not adequate for Census Bureau use. In its surveys and censuses, the Bureau assigns at a minimum 4-digit industry codes based on primary products produced or distributed, or services rendered. There are 1100 such 4-digit SIC codes. In contrast, the Principle Industrial Activity (PIA) code available from IRS covers less than 200 business descriptions at the 2 or 3-digit

level only and is based on taxpayer self-classification for corporations and partnerships, and written self-descriptions for sole proprietorships. The PIA code is used primarily for non-employer businesses identified during each economic census. SSA industry coding is based on the respondent's written description of activity on the EI number application form and encompasses approximately 900 2, 3, and 4-digit codes.

Also, for single establishment companies derived from administrative records, the mailing address rather than actual physical location is sometimes the only geographic information available. The use of mailing addresses to compile geographic statistics obviously can result in biases in the data, particularly for cities and other places in highly urbanized areas.

Sometimes administratively convenient concepts are used that do not always provide the most statistically advantageous data, such as IRS tax reporting on the legal entity or EI basis which often includes combinations of individual establishment data; or SSA's Establishment Reporting Plan which similarly provides for respondent data to be grouped and reported as a single entity for what in fact are two or more establishments. This is particularly troublesome in intercensal years, as establishment detail for new multiestablishment companies is not available to the Directory.

Changes in the administrative requirements of the agency supplying the data can affect Bureau programs. For example, through 1977 the Bureau received edited quarterly payroll data, wage item information and first quarter employment from SSA. The data were used not only as a source of employment, but also as a further refinement in the editing and imputation of IRS payroll data. Starting in 1978, SSA changed to an annual reporting system and the quarterly data were no longer available. IRS agreed to add the employment item to the 941 payroll reports, and the edit and imputation methodologies were redesigned.

Generally, problems such as these occur because statistical uses of the data are secondary to the administrative needs of the particular agency. The advantages of the administrative record system, however, far outweigh the disadvantages and provide the underpinnings that are essential for the central list concept.

III. Processing and Use

The Census Bureau has long recognized the need for a standardized list of firms and establishments for use within the Bureau itself. Experience with the 5-year economic censuses demonstrated that more frequent list updating would be essential. Earlier censuses depended on a separate pre-canvass process to sort out 4 years of mergers, acquisitions, internal reorganizations, births, deaths and other changes in the business universe. During the earlier censuses, the ID numbering system now used in the Directory evolved, in part, to keep track of the constantly changing structure of multiestablishment companies.

It also became evident that a link to the administrative record systems of IRS and SSA would be needed. These EI or legal entity level data would be of enormous assistance in controlling and reconciling the massive coverage operations performed for multiestablishment companies. They

were needed also as proxy data for millions of smaller single establishment companies, which would greatly reduce reporting burden and processing costs.

As a result of the many functions involved in taking the economic censuses, from preparing a mailing list to compiling the results, a centralized file processing and maintenance system was developed. This system, essentially the forerunner of the Directory, would assemble, integrate and control the information needed for Census processing.

In fact, the initial Directory was established in 1973 using the 1972 Census file as the base register. Multiestablishment companies not covered by the economic censuses (for example, companies classified in the finance, insurance and real estate industries) were identified in a supplemental survey conducted concurrently with the censuses. Beginning in 1974 and continuing in subsequent years, an annual Company Organization Survey has been undertaken to ensure that the organizational structure of each multiestablishment company would be updated regularly. This survey includes companies in scope of the economic censuses as well as out-of-scope companies covered in the special survey described above. For companies with 50 or more employees, the survey is mailed annually, while companies with fewer than 50 employees are canvassed once in a 3 year cycle. Directory information for multiestablishment companies is updated with company structure and affiliation changes, name and address changes, new establishments, basic data items, geographic codes, SIC codes, legal form of organization, etc.

Single establishment businesses are identified and updated from administrative record sources. The Business Master File (BMF) of IRS serves as the basic universe file from which single establishment companies are derived. The BMF is a name and address listing of all EI numbers, or legal entities, in the United States. First quarter employment and four quarters of payroll from the IRS form 941 files are then added and used to derive establishment size codes. For agricultural entities, the data come from the IRS 943 files. Files containing SIC codes and geographic information for all business birth EI's are received monthly from the Social Security Administration. IRS administrative record business receipts are also used in the Directory, but only during census years.

This administrative record information, in conjunction with the Company Organization Survey, forms the nucleus of the Directory. Additional information is also fed into the system from each 5-year economic census and from the Bureau's on-going current surveys such as the Annual Survey of Manufactures, the Current Industrial Reports Survey and the Current Business Reports Survey.

Various editing, monitoring and imputation methodologies are applied to the data to ensure completeness and reliability, including: systematic monitoring of the employer universe, as determined by the filing of IRS quarterly 941 employment and payroll reports; editing the data content of the quarterly 941 files for dimensionality and reasonableness; imputation of missing or erroneous payroll or employment entries in the 941 files; evaluating the completeness and coverage of all multiestablishment companies reporting

annually in the Company Organization Survey; imputation of missing data items for multiestablishment companies; editing year to year relationships of data aggregated to the SIC within county levels, in preparation for publishing data in County Business Patterns.

These ongoing Directory operations annually produce a standardized, current, unduplicated list of employer firms and their establishments. As the Directory evolved operationally, its practical use as a universe list became possible and it is used currently at the Census Bureau in many programs including:

- 1977 Economic Censuses - the Directory was used as the basic mailing list for the censuses, replacing the burdensome pre-cavass operation;
- 1978 Census of Agriculture - because the Directory identifies large farm employers and multie-establishment farms, the need for a pre-cavass to identify this portion of the farm universe was unnecessary;
- Annual Survey of Manufactures - the annual mail, check-in, and other processing operations for this survey have been assimilated into Directory operations, effecting material economies of scale and reducing respondent reporting burden;
- Current Industrial Reports - the Directory is being used in several of the current surveys to provide data on new firms not previously included;
- Current Business Reports - the Directory is used as the basic universe from which samples are drawn, new businesses identified, etc.;
- County Business Patterns - through use of the Directory, this program now provides data at the establishment level, rather than the broader reporting unit level, and data on total payroll rather than only payroll subject to social security taxes;
- 1980 Decennial Census - the Directory was used to provide place-of-work coding.

Other uses include reimbursable projects performed by the Bureau for requesting agencies. Without the availability of the Directory, the cost of establishing frames for these surveys would probably have been prohibitive. These projects have included work for: the National Science Foundation, the Export - Import Bank, the Department of Energy, the Council of Economic Advisers, and the Department of Labor.

The development of the Industrial Directory has benefitted the Census Bureau greatly, not only because of cost savings but also because of qualitative improvements. The centralized processing, the standardization of updating and editing systems, the uniform assignment of SIC and geographic codes, the interchange of data and other information among Bureau programs, the feedback of survey results and corrections and more have produced a consistent, highly reliable business universe list. As yet, however, the potential of this list has not been approached.

IV. Future

Although the Directory is operational at the Bureau, it is still far short of its objectives. Reaching these objectives requires that many issues be examined and resolved.

Foremost among these is the current census law, title 13, U.S. Code, under the authority of which

the Bureau conducts its statistical operations, including the gathering of information for the Industrial Directory. The confidentiality provisions of title 13 prevent the Bureau from releasing any such information to other agencies.

Title 26, the Internal Revenue Code, is also a barrier to other agency use of the Directory. The supplying of IRS administrative record data to the Census Bureau is authorized in title 26 and, as a result of the Tax Reform Act of 1976, use of that return information by any other agency is prohibited. These blanket restrictions preclude the release of any Directory information, even the general statistical information contained in the Directory that would not, in fact could not, breach the confidentiality of any firm.

These laws must be amended if the objectives of the central list concept are to be realized. A coordinated interagency effort, under the direction of OMB, is underway to make these needed changes. Legislative amendments to title 13 and title 26 have been prepared. Efforts to introduce this enabling legislation in 1979 and early 1980 were stalled because of the uncertainty surrounding the proposed Reorganization of the Statistical System. The effort has been renewed, however, and additional refinements were made to the bill, reflecting agency comments made during the legislative clearance process.

Under provisions of the bill, conditions for release of Directory information are strictly limited. The Directory can be used for statistical purposes only and is limited to authorized statistical agencies or their units. The information may not be used for purposes of regulation, investigation or enforcement. Only the general information mentioned earlier can be released. Because the Directory contains no quantitative data per se, but rather size and activity codes, release of this information would not compromise the confidentiality of a firm's operations.

In its history, the Census Bureau has never breached the confidentiality of company information. The trust and reliance developed between Census and the business sector are crucial to the collection of timely, accurate statistics. It is important to emphasize that the legislative amendments do not jeopardize our commitment to protect the confidentiality of a firm's information.

After a comprehensive review of the Directory program, the General Accounting Office issued a report to the Congress in 1979. In that report, the Comptroller General strongly endorsed this legislative initiative and recommended that Congress favorably consider the amendments.

Specific guidelines governing use of the Directory will be needed if access is approved. A working group of the Interagency Committee has drafted a set of regulations establishing the criteria for agencies or units to be eligible to use the Directory. The regulations also specify requirements that must be met before Directory information can be released to eligible agencies, such as establishing appropriate administrative, technical and physical safeguards to ensure the security of information released. The design and implementation of this system of safeguards is prerequisite to the Directory becoming an inter-agency tool.

An interagency technical committee has been

established to address these and other issues. The committee will include representatives of the Census Bureau and each potential source or user agency. Meetings have started already with the U.S. Department of Agriculture and the Bureau of Labor Statistics. The objectives of the Committee generally are to identify and resolve operational concerns that would affect interagency use of the Directory. A broad range of topics will require work, including:

- Examining each potential user agency's needs from the Directory and determining the suitability of the Directory to meet those needs. For example, although the establishment may be the ideal reporting unit for presenting basic information on an industry and geographic basis, it is not necessarily the most useful for all reporting purposes. Some agencies, particularly those concerned with financial data, find the establishment concept inappropriate. Where necessary, alternative linking systems must be developed that can accommodate different reporting units, or groupings of establishments, used in some statistical programs;

- Examining each agency's potential updates to the Directory. Before information can be accepted and used to update the Directory, its derivation and reliability must be known. This process will be critical to ensure definitional compatibility and quality. Evaluation and/or reconciliation studies will be necessary to determine the relative quality of information from different sources.

- Developing ranking systems to resolve conflicting information. In any system involving multiagency updates, information may vary for the same establishment, legal entity or company. Edits will be developed not only to test incoming information for completeness and reasonableness, but also to identify conflicting information and to accept or reject that information. This process will depend heavily on the reconciliation studies noted earlier.

- Establishing the systems or procedures for accessing and using the Directory.

- Establishing the systems or procedures for updating the Directory with feedback information from other agencies. Just as is done now with the administrative record information from IRS and SSA, systems will be developed for receiving and processing data from other agencies.

- Developing an update and availability schedule. It is anticipated that the content of the Directory will be changing constantly. These changes must be monitored, coordinated, and even forecast so that all participants can optimally plan and schedule their uses. For example, if the latest name and mailing address information were expected from the IRS Business Master file two weeks after the proposed mailing of a survey, it may be desirable to reschedule.

- Improving the content of the Directory. For example, certain limitations are inherent in the use of employment and payroll data as measures of size. In capital intensive industries, employment or payroll is relatively unimportant in determining economic significance for sampling purposes. Sales and business receipts data on an annual basis would greatly improve utility of the Directory, but will require that certain conceptual issues be resolved including:

- Determining the most meaningful definitions of receipts for different economic activities. Business receipts alone understate significantly the receipts activity of particular industries not within the scope of the economic censuses, such as banking, insurance and real estate. Total receipts, which include other revenue items such as interest income and dividends, are a more complete measure of economic activity for these areas.

- Determining the practicality of reporting receipts data for certain types of companies. Airlines, for example, or banks, may have difficulty reporting receipts data for each physical location operated.

- Determining the compatibility of receipts data available from administrative records with Directory definitions. For example, in wholesale trade the administrative receipts for agents and brokers represent commission income, whereas gross revenues are necessary for census statistics.

Other improvements are being explored also, such as the addition of inventory and capital expenditures data, and the use of product class codes and secondary activity SIC codes to supplement establishment level primary SIC codes. For surveys designed to collect specific product information (rather than establishment level total activity), SIC codes do not always identify the target population. Codes representing product level or secondary activity detail may be incorporated, after an assessment of need and practicality.

In addition to these issues, the committee provides a forum for another of the Bureau's objectives: fostering an understanding of the Directory that we feel is essential to its success. The Directory can be used to its potential only if its characteristics, its profile, its nature, are understood by its users. Without this knowledge, the Directory would be at best underused, and at worst misused.

For the Directory to respond adequately to the exigencies of the modern statistical processing era, a technological upgrading of the system is needed. Currently, the Directory file is recorded on computer tape, supplemented by duplicate paper files or listings. Corrections and changes to the Directory must be accumulated and processed in batches, sequentially. Because of the size of the file, cost considerations make frequent updates impossible. Consequently, separate subsets of the Directory used in mailing the Bureau's current surveys must be maintained to expedite their processing, and throughout all phases of Directory processing, annotated listings must be clerically maintained to keep track of incoming survey information and process reports. This system requires a huge clerical effort, causes processing inefficiencies, and results in a considerable amount of information sitting in the pipeline waiting to update the Directory. The nature of collecting economic information, and efficient use of that information by participating agencies, demands that all file processing be as timely as possible.

The Census Bureau recognized that to fulfill the objectives of the central list concept, and to meet the increasingly complex needs of our own programs, these technological constraints must be

removed. Three, six or sometimes more months can not elapse before basic collected information is incorporated into the Directory. Critical information such as changes in company affiliation or internal organization, changes in address, basic economic data, the addition of new businesses, the identification of out-of-business cases, etc., must be accounted for in the file as quickly as possible.

To meet these objectives, and to meet the massive processing requirements of the economic and agriculture censuses, an on-line, interactive computer system will be implemented in 1982. This system will offer the following capabilities:

- Bar Code Assisted Check-In, Sorting, and Microfilming - these forms processing operations will be automated through the use of bar coded labels containing file numbers and processing codes. Use of optical scanners and hand held wand readers will expedite the check-in and sorting of forms and the recording of microfilm frame reference numbers.

- Combined Data Entry and Screening - unlike previous censuses, processing for 1982 will combine the data entry and screening operations into a single processing step performed on a flow basis. This will permit faster release of the keyed data because the screening for report form problems will be mechanized and will occur simultaneously as data are entered into the computer.

- Centralized Processing System - this computer system will provide further automation through interactive terminal access to the Directory. It will allow inquiry access on a flow basis to determine completeness of responses from individual companies to support editing and correction of data, to assist in resolution of coverage and classification problems, and for a variety of other operations. It also will allow direct access to the Directory for the critical matching of reported data to previously assembled control information, such as from administrative records.

These technological improvements and more are being explored and developed by the Bureau to support a fully functioning Directory. For example, the Bureau will be developing standardized sampling and tabulating packages to be used in concert with the on-line direct access system. User supplied parameters, such as stratification,

sample size, cost, and variance would drive the program. Also, the capability for longitudinal analysis will be developed and built into the system. This capability will include the tracking of activity at individual physical locations over time, irrespective of ownership changes. In this way, the evolution of economic activity in specific local areas can be quantified.

An on-line system and other improvements will allow also for monitoring the status of, and controlling, the inclusion of any given establishment in a survey or census. An establishment will be represented in the Directory only once, irrespective of the number of surveys in which it appears. All changes, regardless of source, will be applied to the same entity. A monitoring system will flag the establishment as ineligible for further surveys if its inclusion rate, or response burden index, exceeds a predetermined limit. This system will also ensure that any access to the Directory is consistent with established policies for protecting against misuse or wrongful disclosure of Directory information.

The advancements in telecommunication alone offer astounding capabilities in data sharing. If adapted to the Directory system, interagency use and interchange of information could be accomplished virtually instantaneously through a shared network of terminals and other peripheral equipment. Currently, information used to update the Directory must be keyed onto magnetic tape and, in the case of IRS administrative record data, those tapes must be delivered to the Census Bureau and read into the computer. These cumbersome steps could be eliminated through use of new technology involving optical character reading and simultaneous data transmission. With a shared direct access network in place, the Directory's reservoir of economic information would be available instantaneously to users. Given that appropriate safeguards exist, analysts, policy and decision makers, survey personnel, etc., throughout the government could query the data base, conduct analytical modeling, select samples, tabulate data sets, and so on. There are many possibilities.

The Census Bureau is making substantive efforts in all these areas. The success of the Directory will depend not only on these efforts, but on the continued cooperation of the business community and an understanding and objective evaluation of the merits of the system by all concerned.