CENSUS BUREAU USES OF ADMINISTRATIVE DATA FOR SERVICE STATISTICS

Howard N. Hamilton, Paul S. Hanczaryk
Brian V. Greenberg and Carl A. Konschnik

United States Bureau of the Census
Washington, DC 20233

To Be Presented At

Sixth Meeting of the Voorburg Group
On Service Statistics

Helsinki

October 7-11, 1991

CENSUS BUREAU USES OF ADMINISTRATIVE
DATA FOR SERVICE STATISTICS

Howard N. Hamilton, Paul S. Hanczaryk,
Brian V. Greenberg and Carl A. Konschnik
Bureau of the Census

I.   Introduction--Sources and Uses of Administrative Records

The United States statistical structure is highly decentralized with
different agencies assuming differing responsibilities. In order to
make efficient use of data already collected and to reduce respondent
burden, cooperative arrangements have been established between agencies
to allow for transfer of information, often administrative records.
Such information transfers are very specific and well-defined and
typically embedded in the regulations under which agencies operate.

Administrative records data used by the Census Bureau are obtained
chiefly from the Internal Revenue Service (IRS) and to a lesser extent
from the Social Security Administration (SSA). Each of these agencies
has close and ongoing contact with the business community. Such
administrative data play a major role in the entire economic statistics
program at the Bureau of the Census. Through the administrative tax
system the Census Bureau obtains virtually a complete list of businesses
engaged in all types of economic activity and selected data items for
each business. These data allow for the publication of a variety of
statistics at a comparatively low cost and with much less burden on
responding business firms.

The IRS and SSA information is obtained by the Census Bureau under
provisions of the Internal Revenue Code as amended by the Tax Reform Act
of 1976. Before that, IRS records were available through a series of
Executive orders while SSA information was obtained through an
administrative order of that agency. Information from administrative
records is compiled under the confidentiality provisions of the law
under which the Bureau operates (Title 13 U.S.Code), as well as under
the confidentiality requirements of each supplying agency. Title 13
prohibits the Census Bureau from releasing any data that can directly or
indirectly be used for purposes of taxation, regulation, or
investigation.

Since the 1954 Economic Censuses, the Census Bureau has used
administrative records extensively to assemble mailing lists and obtain
industry classifications for businesses included in the economic
censuses and subsequent surveys.

For processing purposes, the Census Bureau differentiates between single-establishment and multi-establishment firms. The administrative data information obtained annually for single-establishment businesses includes four quarters of payroll, March 12 employment, name and address, Employer Identification Number (the taxpayer identifier for businesses with employees), legal form of organization code, and the industrial activity code.

Administrative records for multi-establishment firms typically provide data for entire companies or other legal entities. Thus they do not provide information at the establishment level (physical location) that is required for census or survey activities. The Bureau conducts the annual Company Organization Survey to collect information on the individual establishments of multi-establishment companies.

Information from the administrative system for both single-establishment and multi-establishment companies provide the framework that is essential for the Census Bureau to construct and maintain its central business list—the Standard Statistical Establishment List (SSEL). The SSEL is a data base of United States businesses and their establishments compiled from administrative data files and responses from the Company Organization Survey and other Census Bureau data collection vehicles. It contains basic economic information for some 6.8 million corporations, partnerships, individual proprietorships and other businesses with employees. Information for another 7.5 million nonfarm businesses without employees is compiled once every 5 years at the time of the economic censuses.

Due to costs, the Census Bureau has not maintained an annual up-to-date ongoing listing of nonemployers on the SSEL because they are typically small businesses operated by individual proprietors or working partners. They contribute less to Census Bureau programs than employer businesses —accounting for less than 2 percent of business receipts for the total of both the goods producing and service sectors. This percentage is higher in the services industries and is large for certain kinds of business activity. Later in this report we discuss plans to maintain an annual list of nonemployers to use in all Census Bureau programs with an emphasis on impact in the service statistics program.

In Section II we provide examples of ways administrative records are employed in a variety of settings. One can easily see how pervasive the use of administrative records is. In Section III there is a discussion of cooperative arrangements between agencies to expedite the flow of administrative records and other interagency data transfers. In Section IV we describe current planning efforts to expand the use of administrative records for both current surveys and censuses. Much of the discussion in this report focuses on the uses of administrative records across many sectors of the economy. In Section V the focus is on service industries and we further describe the impact of administrative records in that area.

II. **Description of the Principal Administrative Record Files Used for the Economic Statistics Programs at the Census Bureau**

In this section we describe the principal administrative record files received from the IRS and SSA along with their uses. Most estimates of numbers of records and uses described are based on the 1987 Economic Censuses and related activities.

A large part of the procurement costs for administrative record files is fixed (i.e., file extraction, disclosure review, coordination, etc.). Accordingly, the marginal cost of obtaining an additional administrative record data item is small if the item is keyed as part of normal IRS processing. Substantial costs do result, however, from specially keyed data items, that is, items not keyed as part of the normal processing, but keyed specifically for Census Bureau programs.

1. **Business Master File (BMF)**

The Bureau of the Census receives an annual extract containing over 13 million records from the Internal Revenue Service's Business Master File. The Business Master File (BMF) is a complete list of employer businesses in the United States. In addition, monthly BMF supplement files are received that detail current changes to the annual extract. Name and address information for all single-establishment firms on the SSEL is obtained from the BMF. We also use tax filing requirement codes, industrial activity codes, and employment codes from the BMF in current programs. Census Bureau uses of various data item are detailed in Attachment 1, Table A.

2. **Form 941 Series and Form 943 Payroll and Employment Files**

The Census Bureau obtains employment and payroll data from IRS Forms 941 (Employer's Quarterly Federal Tax Return) and 943 (Employer's Annual Tax Return for Agricultural Employees). The 941 file defines the population of, and provides basic data for single-establishment firms on the SSEL. Form 941 data are also used annually in an edit that evaluates the coverage and completeness for all multi-establishment companies and are used each census to determine the single-establishment mail panel and provide proxy data for nonmail businesses. Over 27 million quarterly 941 records are received each year from the IRS. For agricultural employers, March 12 employment and annual payroll data come from the 943 files. See Attachment 1, Table B.

3.   <u>1985 Business Income Tax Return Files</u>

We received selected 1985 business income tax data from
Forms 1120; 1120-A; 1120S; 1065; and 1040, Schedule F. The
IRS provided tax return information only for those
businesses with indications of agricultural activity. These
data were used to develop the mailing lists for the Farm and
Ranch Identification Survey and the 1987 Census of
Agriculture. See Attachment 1, Tables C1 and C2.

4.   <u>1986 Business Income Tax Return Files</u>

We used selected 1986 business income tax data from Forms
1120; 1120-A; 1120F; 1120S; 1065; and 1040, Schedules C and
F in the 1987 Economic and Agriculture Censuses. Tax
records with indications of agricultural activity were used
to develop the mailing lists for the 1987 Census of
Agriculture. The tax files provide a source of industry
classification (for County Business Patterns as well as for
the economic censuses), particularly for sole
proprietorships. The receipts data provide a source of
prior year receipts for the censuses, and may be used as an
editing or imputation device when 1987 administrative
receipts data are missing or in error. See Attachment 1,
Tables D1 and D2.

5.   <u>1987 Business Income Tax Return Files</u>

The IRS furnished selected 1987 business income tax data
from Forms 1120-A; 1120F; 1120S; 1065; 990; 990C; 990PF;
990T; and 1040, Schedules C and F. These data were used in
the 1987 Economic and Agriculture Censuses to: provide
receipts data and other information for nonmail and
nonresponse businesses; define the potential economic
censuses nonemployer universe (i.e., if receipts data
present, but no matching payroll record); furnish addresses
used for geographic coding; provide end-of-year and months-
in-business information; and form the initial universe for
the Surveys of Minority-Owned and Women-Owned Businesses.
Data items from the 1987 tax files are used in the sample
selection and control of the various current surveys of the
Census Bureau economic divisions. See Attachment 1, Tables
E1 and E2.

6.   <u>Quarterly Financial Report (QFR) Sample File</u>

We obtain an annual extract of information from IRS for a
sample of corporations on the BMF. Information on
corporations are used in the preparation and processing of
the QFR publications. See Attachment 1, Table F.

7. Form 851 Affiliations Schedules File

We obtained parent/subsidiary information from Form 851 from IRS for tax years 1982 and 1987. Preparation of the economic censuses reports entitled "Link of Census Establishments and IRS Corporation Data" depends on our access to the Affiliations Schedules. These data were used to identify the subsidiary corporations of parent corporations that file consolidated returns. See Attachment, 1, Table G.

8. SSA Birth File

We receive selected birth information from Form SS-4 (Application for Employer Identification Number) from SSA. Standard Industrial Classification codes from the SSA birth file are the primary source of classifications for new single-establishment firms on the Standard Statistical Establishment List. The SSA birth information is also used to identify potential births for the Quarterly Financial Report program. See Attachment 2, Table A.

9. Self-Employment Tax File

We obtained from SSA selected information from Form 1040, Schedule SE for tax year 1987. These data were used to determine sole-proprietor ownership for the Minority-Owned and Women-Owned Businesses programs. The 1982 self-employment tax file contained over 8 million records. See Attachment 2, Table B.

10. Race and Sex Coding File

The Census Bureau furnished SSA with over 15 million Social Security Numbers (SSNs) from a file of 1987 businesses. The SSA provided race and sex codes for use in our Minority-Owned and Women-Owned Businesses programs. See Attachment 2, Table C.

III. Census Bureau Cooperative Arrangements with Other Agencies

A. Procedure for Obtaining Administrative Records

The statute which authorizes the Census Bureau to obtain tax return information from IRS is Section 6103(j)(1) of the Internal Revenue Code of 1954. Specifically, it appears as below:

(j) Statistical Use--

    (1) Department of Commerce--upon requesting in writing by the Secretary of Commerce, the Secretary shall furnish

        (A) such returns, or return information reflected thereon, to officers and employees of the Bureau of the Census, and

        (B) such return information reflected on returns of corporations to officers and employees of the Bureau of Economic Analysis,

        as the Secretary may prescribe by regulation for the purpose of, but only to the extent necessary in, the structuring of censuses and national economic accounts and conducting related statistical activities authorized by law.

The specific data items to be received by the Census Bureau are listed in the Regulations on Procedures and Administration of the Internal Revenue Service. The regulations also allow specific tax return data maintained by SSA to also be given to the Census Bureau.

Regulations relating to Section 6103 must be periodically modified to permit access to specific data elements. Each year the Census Bureau must justify its specific needs and uses of the data and detail disclosure and security measures. Before the beginning of each fiscal year, the Bureau is required to issue a report to IRS detailing intended uses, retention policy and safeguards of IRS data for that year.

It takes over two years to incorporate changes (usually to add data elements) to the IRS regulations. A formal data request (for data authorized by the regulations) requires anywhere from 9 to 18 months before the data are received. This time is used by IRS for disclosure review (to verify that the requests are covered by the regulations) and to prepare their computer programs and system to do the required extracting of data.

B. **Liaison Personnel for Administrative Records**

Recognizing the vast importance of administrative record data, the Census Bureau has set up specific positions which serve as pivotal links to the IRS and SSA. The individuals with these responsibilities work closely and cooperatively with the other agencies in ensuring good communication and smooth administrative record data flow. Regular and special meetings are held between agencies where issues and needed changes are discussed and resolved. These key people play a major role in providing the administrative record base for the statistical programs of the Census Bureau.

Interactions regarding administrative data uses and their acquisitions are not static; they have evolved, and will continue to evolve, over time.

C. **Quality Assurance of Administrative Data**

Over the past several years, the quality of the administrative records generally has been excellent. This has not always been the case, however. Serious problems were noted with the quality of the 1982 business income tax data from the IRS. This showed up chiefly as inadequate quality of the principal industrial activity codes for sole proprietorships. As a result of this problem, the Census Bureau published only limited statistics for nonemployers in the 1982 Economic Censuses. If the quality assurance programs had been more sophisticated, the errors could have been identified earlier and the negative impact would have been minimized.

Heading into the 1987 Economic Censuses, it was determined that additional measures were needed to ensure the quality of administrative record data received from the IRS. An overall quality management system was necessary. The three major factors that have plagued us in the past are:

1. **Vast Amounts of Administrative Record Data**

    The IRS provides selected business 1987 tax return data (received in 1988) for various legal forms of businesses, including corporations, S corporations, foreign corporations, partnerships, nonprofit organizations, and sole proprietorships. In total, the Census Bureau received about 75 million tax return records in 1988. Table 1 details the approximate number of administrative records used in the 1987 Economic and Agriculture Censuses for the various form types. Clearly, the number of data records received during census years is immense, but the complexity of the required quality assurance goes beyond sheer volume. A data record often contains several data items, each greatly increasing the detail of the individual records and the entire data files. Moreover, not all form types contain

the same set of data items, nor do they have the same pattern of receipt. Consequently, in addition to performing quality review for about 75 million individual records, the Census Bureau must also be concerned with assuring the quality of the various data items on those 75 million records.

Businesses file their tax returns with one of ten IRS Service Centers. Each of the individual centers processes the returns, and the quality of data received from different service centers can vary. The Census Bureau reviews data at the service center level in response to such variation.

2. Restrictive Resources

Restrictive resources are another major factor that contribute to the difficulty of assuring the quality of the administrative record data. In keeping with the overall governmental policy on spending, the Census Bureau is striving to provide greater services at less cost. Workloads for programming staffs increase significantly during census years, yet the staffs do not expand proportionately. The quality assurance processing, which relies considerably on various computer resources, can be adversely affected.

3. Lack of Communication Between Agencies

Miscommunication or lack of communication between agencies has contributed to past administrative record problems. Clear lines of communication between the Census Bureau and the agency providing the data during all phases of the procurement process also are essential for assured data quality. The agencies first must agree upon the data files and the specific data items that are needed and that can be provided. Certain data that the Census Bureau requests may not be available or in some cases affordable. Any discrepancies must be resolved in time to avoid delays, which could affect data utility. Moreover, the agencies must agree upon the expected quantity and quality of the administrative data. Requirements that quantify the Census Bureau's expectations of the incoming data must be established.

The development and implementation of the quality assurance system represents a comprehensive response to the administrative record data problems encountered in the past. The system provides for the review of large and complex IRS data files, promotes frequent interagency communication, and identifies errors on a timely basis. Basically, the Census Bureau establishes standards that detail the fundamental requirements of the incoming IRS data and patterns of the actual data are compared to these standards. The

Census Bureau then prepares status reports for IRS indicating whether the data conform to the standards.

Census Bureau staff develops standards far in advance of the actual receipt of the data and shares this information with IRS staff. This gives the IRS ample opportunity to examine the requirements for reasonableness and negotiate adjustments if necessary. The requirements cover both timing standards and quality standards. The timing standards list the estimated total number of tax returns for the different types of businesses and the estimated number to be received by various dates. The quality standards delimit the expected reporting patterns of specific data items.

A series of data verification results tables are created that compare the actual data to the expected standards. Discrepancy flags are set for those data components that do not meet the standards. This approach minimizes the risk of analytical omissions during the review process.

Status reports comparing the reporting patterns of actual data to the pre-determined standards are sent to the IRS monthly and promote communication between the agencies. These reports are a subset of the comprehensive results tables, detailing only the basic requirements of the IRS data set. If data problems exist, they are illustrated in the report. Immediately, the Census Bureau and the IRS must decide upon any remedial action or recovery efforts necessary to prevent compromising the censuses. Timeliness is crucial because the IRS data tapes are not kept indefinitely. If errors are not identified early and remedial action is not implemented in time, recovery of the data may not be possible or may become extremely costly.

The quality assurance system is not designed to guarantee that administrative data problems will never occur. It does serve, however, to document requirements formally so that the characteristics of the data set are not left to chance, and monitoring and early error identification are possible. Briefly, the predefined standards establish to limits of tolerance.

TABLE 1

The Approximate Number of Administrative Records Used in the 1987
Economic and Agriculture Censuses for the Various
Form Types by Tax Year

Type of Record	Number of Records		
	1985	1986	1987
Business Income Tax Files	2,617,000	20,051,000	30,881,000
Form 1040, Schedule C	—	11,750,000	12,500,000
Form 1040, Schedule F	2,450,000	2,450,000	—
Form 1040, Schedule SE	—	—	10,000,000
Form 1120	42,000	2,550,000	2,650,000
Form 1120-A	—	200,000	210,000
Form 1120F	—	11,000	11,000
Form 1120S	17,000	900,000	950,000
Form 1065	108,000	1,750,000	1,800,000
Form 990	—	380,000	400,000
Form 990-PF	—	35,000	35,000
Form 990-T	—	25,000	25,000
Form 1120S, Schedule K-1	—	—	700,000
Form 1065, Schedule K-1	—	—	1,600,000
Annual Tax Files	41,950,000	43,500,000	45,050,000
IRS Business Master File	24,000,000	25,000,000	26,000,000
IRS Payroll & Employm. File	17,000,000	17,500,000	18,000,000
SSA Business Birth File	950,000	1,000,000	1,050,000
Total	44,567,000	63,551,000	75,931,000

D.   Initiative to Access Firm Data for the Bureau of Labor Statistics

The Office of Management and Budget (OMB), charged with the
responsibility to review and approve proposed report forms for the
various statistical programs, has been encouraging agencies such
as the Census Bureau and the Bureau of Labor Statistics (BLS) to
share information on businesses, where this is permissible by law.
In a recent memorandum of understanding between the Bureau of the
Census, U.S. Department of Commerce and the Bureau of Labor
Statistics, U.S. Department of Labor the following purpose was
stated:

This agreement establishes the procedures under which the Bureau of Labor Statistics and the Bureau of the Census will implement the Office of Management and Budget order of December 20, 1990. That order determined conditions necessary for:

1) Bureau of Labor Statistics access to specific tax information and Title 13 data; and

2) Bureau of the Census access to certain Standard Industrial Classification (SIC) information collected under the sponsorship of the Bureau of Labor Statistics;

for the purpose of assisting the Bureau of the Census to improve the coverage and accuracy of its business classifications. The Bureau of the Census will use SIC information provided by the Bureau of Labor Statistics to the maximum extent feasible in lieu of conducting direct inquiries of business respondents. In addition to this primary purpose, the activities covered by this agreement will also support a joint assessment of matching operations, including evaluation of the quality of information produced, any practical problems, and the cost-effectiveness of an ongoing matching program.

At present, the Census Bureau is reviewing the results of the first of several groups of matches to the Bureau of Labor Statistic's Business Establishment List (BEL) to determine to what extent current and reliable SIC information can be picked up from the BEL. A report on the results of this investigation is due to OMB by the end of 1991. These results will determine whether a program of using BEL supplied SIC codes would be beneficial to the Census Bureau's Statistical programs.

IV. Current Initiative to Expand the Use of Administrative Records and Improve Matching to the SSEL

Administrative records are used for tabulation in the censuses for nonemployers and for employer businesses which do not receive a mail form or which are nonrespondents. Currently the receipts files are obtained only in the census year and the year prior to the census. These annual receipts data have not been available for use in annual or subannual surveys, either for employer or nonemployer businesses. As part of our ongoing data improvement program through use of administrative records the Census Bureau plans to request these receipts data on a regular annual basis from IRS.

An important capability in administrative record processing is to be able to link the receipts data for business taxpayers with their payroll data on the SSEL. In the past we have used only the EI number as reported on the 1040-C form (with an SSN as its primary identifier) to link a sole proprietorship's 1040-C tax return to its payroll based record (with an EI as its primary identifier) on the SSEL. Because employers often fail to include their EI number on the 1040-C, the necessary link is often not made. In addition to using the EI from the 1040-C, the SSN and other variables on the BMF, including names and addresses, can be used to link the 1040-C to its associated EI on the SSEL. Some preliminary studies have shown that a sole proprietorship EI with payroll on the SSEL but with no receipts (because of failure to match to its 1040-C record) could often to be linked via a name and address match, using both the primary and secondary names and the address from the BMF. More investigation is necessary to determine the most efficient method for this name and address match.

Potential benefits to be derived from the annual use of tax receipts data and better linking to payroll records on the SSEL are:

1. Receipts data will be more complete for the census year (due to improved linking of the EI on the SSEL and the SSN based 1040-C files) allowing more census nonmail or nonresponse cases to use administrative data rather than some value imputed by other means for census tabulation.

2. The determination and tabulation of census nonemployers will be improved by removing much of the employer duplication which results on the 1040-C file due to failing to make the proper link to its EI on the SSEL.

3. Administrative receipts for survey nonrespondents will be available for use in the annual surveys of Business Division-- which processes the retail, wholesale and services component of the service sector.

4. Business Division will consider using administrative records to tabulate nonemployers in its annual surveys.

5. Business Division will consider replacing its area sample (which represents birth employers and nonemployers in the monthly retail and annual retail and services surveys) with administrative records based data or surveys drawn from the improved and annually available nonemployer lists.

6. Both the census and the current surveys processing will benefit from having similar administrative records processing operations. This will achieve more consistent and improved results for censuses and surveys.

7.   The availability of the Principal Business Activity (PBA) codes provided by the IRS for sole proprietorships will be used to assign SICs and thereby lessen the number of unclassifieds on the SSEL.

8.   Other major programs, such as the County Business Patterns, will benefit from having receipts data for employers and nonemployers available annually, allowing publication expansion possibly.

9.   Improved measure of size computation for business sample revisions will be available. Also, another edit resource for analysts will be provided.

Because of these important potential benefits, the Census Bureau is proceeding to affect these changes at the earliest possible date.

It is worth noting that the proposed changes effect both annual programs and censuses. The impact for economic census programs will be through better linkages between administrative records and the SSEL. This will provide better estimates for non-mail cases in the economic censuses.

For annual programs, the use of administrative records will provide a far more accurate measure for nonemployers. Under current procedures, estimates for nonemployers are derived from an area sample. Every five years, these estimates are benchmarked to the economic census in which nonemployers are estimated through the use of administrative records. By using administrative records on annual surveys to provide estimates for nonemployers, we anticipate a marked reduction in the revision due to benchmarking. In addition, we anticipate a more precise estimate of the nonemployer contribution to estimates in the service sector.

In addition, a general and overall improvement will accrue to our monthly surveys as they are based, in several ways, on the census collections.

V.   Impact on Service Industries of Expanded Use of Administrative Data

Firms in the 1987 Census of Service Industries were divided into the nonmail universe and mail universe. The nonmail universe consisted of firms which were not required to file a regular census return and included:

A.   All taxable nonemployers, i.e., all firms subject to Federal income tax with no paid employees during 1987. Receipts information for these firms was obtained from administrative records. Although consisting of many firms, nonemployers accounted for less than 10-percent of total receipts of all establishments covered in the census. The census included only those service nonemployer firms which reported a receipts volume of $1,000 or more during 1987. Also included were nonemployer firms in operation for less than the full year that reported receipts which, if projected to an annual basis, would have

reached a total of $1,000 or more. Establishments exempt from Federal income tax with no paid employees were excluded as in previous censuses.

B. Selected small employers, i.e., single-establishment firms with payroll below a specified cutoff. Although the payroll cutoff varied by kind of business, small employers generally included firms with 1 to 4 employees and represented about 10-percent of total receipts of establishments covered in the census. Data on receipts, payroll, and employment for employer firms below the payroll cutoff were derived or estimated from administrative records, except for a 10-percent sample of small employer firms in classifications for which specialized data requirements precluded reliance solely on administrative records sources. This sample was included in the mail universe.

The mail universe consisted of firms for which information was obtained by means of a mail canvass and included:

A. Large employers (employer firms above the size cutoff referred to above). Within this category, a report of company organization is conducted periodically to identify establishment locations for firms which operated at more than one location and to obtain information on payroll and mid-March employment at each location. The 1986 Report of Company Organization was used as a coverage check in the census. In the census mailing package containing the appropriate 1987 questionnaires, multiestablishment firms also were sent a form to use in updating the list of establishments.

B. The 10-percent sample of small employer firms in classifications for which specialized data precluded reliance solely on administrative records sources. These firms were sent the census mailing packages containing the appropriate 1987 questionnaire.

The service industries classifications for all establishments were based on the Standard Industrial Classification (SIC) Manual. However, the method of assigning these classifications, and the level of detail at which establishments were classified, differed between the nonmail and mail universes.

For the nonmail universe nonemployers were classified on the basis of information obtained from the administrative records. Selected small employers were classified on the basis of the most current census kind-of-business classification available from one of the Bureau's current sample surveys or the 1982 census. Otherwise, the classification was obtained from the administrative records. If the census or administrative record classifications proved inadequate (none corresponded to a 1987 census classification in the detail required for employers), the firm was sent a brief inquiry requesting information

necessary to assign a 1987 census kind-of-business code. Establishments in the mail universe were classified on the basis of their self-designation, answers to questions on receipts by type of service provided and other special inquiries.

Administrative records are key components of the economic censuses especially retail and services industries. Based on administrative payroll cutoffs, single establishment employer businesses are either mailed a census form or not. For those not mailed a census form, the administrative annual sales or receipts, first quarter payroll, annual payroll and March 12 employment are used for census tabulation. The SIC used for tabulation may also be based on an administrative source. In fact, the economic census of retail trade and of service industries each produce publications of statistics for nonemployers. The data in these publications are based primarily on administrative records.

Nonemployer tabulations for retail and services are based entirely on the Form 1040, Schedule C (sole proprietorship) reports along with (to a lesser extent) Form 1120 (nonemployer corporations) and Form 1065 (nonemployer partnership) tax returns. For the 1987 census for retail (comprised of SICs with first two digits in the 52 to 59 range) nonemployers contributed $47 billion or 3% of the annual sales volume, and about 900,000 or 38% of the number of retail establishments. For the 1987 census for services, (restricted to those SICs covered in the census—namely those with SICs beginning with 7 and most of those beginning with 8) nonemployers contributed $96 billion in receipts or 11% of receipts and about 4.4 million establishments or about 74% of the number of services establishments. These figures indicate the critical role of administrative records in census processing for nonemployers. Due to the forthcoming of expanded coverage in the service industries, we anticipate even greater benefits from our use of administrative records during the 1992 Economic Censuses.

The expanded use of administrative data for the economic censuses and annual surveys will have benefits in all trade areas. However we anticipate more extensive benefits in the service industries than in most other sectors of the economy. For one thing, there are a higher proportion of nonemployers and small employers in the service industries than in other trade areas. And these nonemployers and small employers contribute a greater proportion to estimates of receipts than is the case in other trade areas. As noted above, it is the nonemployers and small employers that are tabulated in census programs with administrative records.

Under current practices, estimates for nonemployers in retail and services are derived from an area sample. Retail stores are generally located in areas where people shop and are typically very noticeable (in order to attract customers). In contrast, establishments in the service industry—especially smaller establishments—may not be located in a central commercial district and may fail to be very visible to the public. In particular, they may easily go undetected in the area sample canvassing. Thus, estimates for small service establishments in the

service industries are probably understated, and we anticipate a marked improvement in estimates for these businesses through an expanded use of administrative records in our annual survey programs.

References

Sections of the following references were used liberally in composing this paper.

Blum, Lawrence A., Bowman, James E. and Sullivan, John M., "Conducting the Economic Censuses," Government Information Quarterly Vol. 4, No. 3 (1987), pp. 263-302.

Bryant, Barbara E. and Norwood, Janet L., "Memorandum of Understanding Between The Bureau of the Census and the Bureau of Labor Statistics," (April 18, 1991). Unpublished Memorandum.

Bugenhagen, Roger H. and Hanczaryk, Paul S., "1987 Economic and Agriculture Censuses: 1987 Business Income Tax File Requirements," (July 22,1986). Unpublished Memorandum.

Complete Internal Revenue Code of 1954, (1985), January 1, 1985 Edition, Prentice Hall, Inc., Englewood Cliffs, New Jersey 07632.

Farrell, Michael G. and Sullivan, John M. (1981), "Industrial Directory - Status and Direction," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 595-600.

Hamilton, Howard N. and Konschnik, Carl A., "Proposal to Update the SSEL with 1040-C and other Administrative Data on an Annual Basis," (July 17, 1991). Unpublished Memorandum.

Hanczaryk, Paul S. and Jonas, James R., "Automated Quality Assurance Processing of Administrative Record Files," Survey Methodology, June 1989, Vol. 15, No. 1, pp. 119-127. Statistics Canada.

Richardson, W. Joel, Hanczaryk, Paul S., Jonas, James R., and Sullivan, John M., "Administrative Record Files Used in the Economic Area," (February 28, 1986). Unpublished Memorandum.

Richardson, W. Joel and Hanczaryk, Paul S., "Industrial Activity Coding of Nonemployer Businesses in the 1987 Economic Censuses," (1988), Third International Roundtable on Business Survey Frames, Auckland, New Zealand.

Trager, Mitchell L. and Hanczaryk, Paul S., "Improving the Quality of Standard Statistical Establishment List Data," Proceedings of Statistics Canada Symposium 90, Measurement and Improvement of Data Quality (October 1990).

U.S. Bureau of the Census, History of the 1987 Economic Censuses. (Draft). To be issued in 1992.

U.S. Bureau of the Census, Census of Service Industries (1987). Nonemployer Statistic Series.

Table A--BMF Data Items and Their Uses
Estimated Procurement Cost: $27,500
Approximate Number of Records Annually: 24 Million

Data Items	Data Item Uses	Special Keyed Items 4/
1. EIN	Provides identification/link to EI records, including SSEL.	
2. Name (first, second, name control, and character counts)	For census and survey mailing purposes. Also used for the Standard Industrial Classification (SIC) name-coding process.	
3. Address (street address, city, state abbreviation, ZIP Code, and character counts)	For census and survey mailing purposes. Also used to assign geographic codes for CBP, other annual economic programs, and economic censuses programs.	
4. Second Reference Code	Not used.	
5. PIA Code	Source of SIC code for the SSEL when other sources are not available.	
6. Transaction Change Code	Denotes the type of change on the BMF. Used for unduplication purposes.	
7. Employment Code	Indicates type of employer. Used to exclude certain employers (e.g., governments) from economic censuses and CBP coverage.	
8. BMF Cycle	Indicates the year and week an update was made to the BMF. Used for unduplication purposes.	
9. Filing Requirement Codes (941, 1120, 1065, 942, 943, 1041, CT-1, 940, 720, 990, 990C, 990T, and 990PF)	Determine legal form of organization for single-establishment firms. Also, active 941 codes identify potential universe for current business surveys.	
10. Service Center, District Office, and Area Codes	Determine service center distributions for quality control purposes. Also used for indexing BMF microfilm.	

4/ The BMF contains no special keyed data items.

Table B--941/943 Payroll and Employment File Data Items and Their Uses
Estimated Procurement Cost: $69,500
Approximate Number of Records Annually: 17.5 Million

Data Items	Data Item Uses	Special Keyed Items 5/
Data Items Common to All Form Types (Forms 941, 941E, 941PR, 941SS, and 943):		
1. EIN	Provides identification/link to EI records, including SSEL.	
2. Master File Tax Account Number	Indicates type of 941/943 return.	
3. Taxable Period Covered	Indicates quarter of payroll for posting to the SSEL.	
4. Employment Code	Indicates type of employer.	
5. Document Locator Number	Distinguishes form type (e.g., 941 and 941E).	
6. First Quarter Employment	Supplies the SSEL with employment data. Provides proxy data for economic censuses nonmail universe.	$30,000
Additional Data Items (Applicable Form Types are Specified):		
1. Total Compensation (Forms 941 and 941E)	Source of quarterly payroll data for the SSEL. Also used in selecting mail panel for economic censuses, and as proxy data for census nonmail universe.	
2. Total Taxable Wages (FICA) (Forms 941, 941PR, 941SS, and 943)	- Forms 941, 941PR, and 941SS--To edit/impute total compensation for the SSEL when combined with taxable tips. - Form 943--Provides annual payroll for agricultural employers.	
3. Total Taxable Tips (FICA) (Forms 941, 941PR, and 941SS)	To edit/impute total compensation for the SSEL when combined with FICA wages.	

5/ The costs of special keyed items are detailed separately.

The 1985 Business Income Tax Return Files (Tables C1 and C2)
Estimated Procurement Cost: $110,000
Approximate Number of Records: 2.5 Million

Table C1--1985 1040, Schedule F Business Income Tax Return File Data Items and Their Uses

Data Items	Data Item Uses	Special Keyed Items 6/
1. SSN	Source of identification.	
2. Validity Digit	Identifies invalid SSNs to be dropped from processing.	
3. Spouse SSN	Source of identification.	
4. Accounting Period	Identifies invalid tax years to be dropped from processing.	
5. Name (first, second, and character counts)	Provides name for Agriculture Division mailing purposes.	
6. Address (street address, city, state abbreviation, ZIP Code, and character counts)	Provides mailing address for Agriculture Division.	
7. EIN	Provides identification/link to EI records, including SSEL.	
8. Gross Profits--Accrual 9. Gross Profits--Cash	Determines size codes for the Agricultural Census mailout.	

6/ The 1985 1040, Schedule F Business Income Tax Return File contains no special keyed items.

Table C2--1985 1120/1065 Business Income Tax Return File Data Items and Their Uses

Data Items	Data Item Uses	Special Keyed Items 7/
Data Items Common to All Form Types (Forms 1120, 1120-A, 1120S, and 1065):		
1. EIN	Provides identification/link to EI records, including SSEL.	
2. Accounting Period	Identifies invalid tax years to be dropped from processing.	
3. Form Code	Distinguishes form types (e.g., 1120 and 1120S).	
4. PIA Code	Source of industrial classification.	
5. Net Receipts	Determines size codes for the Agricultural Census mailout.	

7/ The 1985 1120/1065 Business Income Tax Return File contains no special keyed data items.

The 1986 Business Income Tax Return Files (Tables D1 and D2)
Estimated Procurement Cost: $130,000
Approximate Number of Records: 20 Million

Table D1--1986 1040, Schedules C and F Business Income Tax Return File Data Items and Their Uses

Data Items	Data Item Uses	Special Keyed Items B/
Data Items Common to All Form Types (Form 1040, Schedules C and F):		
1. SSN	Source of identification.	
2. Validity Digit	Identifies invalid SSNs to be dropped from processing.	
3. Spouse SSN	Source of identification.	
4. Accounting Period	Identifies invalid tax periods to be dropped from processing.	
5. Name (first, second, and character counts)	Provides name for Agriculture Division mailing purposes.	
6. Address (street address, city, state abbreviation, ZIP Code, and character counts)	Provides mailing address for Agriculture Division.	
7. EIN	Provides identification/link to EI records, including SSEL.	
Additional Data Items (Applicable Form Types are Specified):		
1. Gross Profits--Accrual (Form 1040, Schedule F)	Determines size codes for the Agriculture Census mailout.	
2. Gross Profits--Cash (Form 1040, Schedule F)		
3. Gross Receipts or Sales (Form 1040, Schedule C)	Source of prior year data. Combined to form net receipts, which is used to edit/impute when 1987 Economic Censuses receipts are missing or in error. For Agriculture Censuses, to develop size codes. (We consider the Surveys of Minority-Owned and Women-Owned Businesses programs to be part of the economic censuses, and have not detailed their uses separately.)	
4. Returns and Allowances (Form 1040, Schedule C)		
5. PIA Code (Form 1040, Schedule C)	Source of industrial classification for the SSEL when other sources are not available. Also, will provide codes for active nonemployer firms in 1987 whose classifications are missing.	

B/ The 1986 1040, Schedules C and F Business Income Tax Return File will contain no special keyed data items.

Table D2--1986 1120/1065 Business Income Tax Return File Data Items and Their Uses

Data Items	Data Item Uses	Special Keyed Items 9/
Data Items Common to All Form Types (Forms 1120, 1120-A, 1120F, 1120S, and 1065):		
1. EIN	Provides identification/link to EI records, including SSEL.	
2. Accounting Period	Identifies invalid tax years to be dropped from processing.	
3. Form Code	Distinguishes form types.	
4. PIA Code	Source of industrial classification for the SSEL when other sources are not available.	
5. Net Receipts	Source of prior year data. Used to edit/impute when 1987 Economic Censuses receipts are missing or in error.	
Additional Data Items (Applicable Form Types are Specified):		
1. Gross Royalties (Forms 1120 and 1120-A)	For selected kinds of businesses, royalties included as receipts.	
2. Net Income (Loss) From Royalties (Form 1065)	For selected kinds of businesses, royalties included as receipts.	
3. Net Farm Profit (Loss) (Form 1065)	Identifies potential farm operators.	

9/ The 1986 1120/1065 Business Income Tax Return File will contain no special keyed data items.

Table E1--1987 1040, Schedules C and F Business Income Tax Return File Data Items and Their Uses

Data Items	Data Item Uses	Special Keyed Items 10/
Data Items Common to All Form types (Forms 1040, Schedules C and F):		
1. SSN	Source of identification for economic censuses purposes (including uses in Surveys of Minority-Owned and Women-Owned Businesses programs).	
2. Validity Digit	Identifies invalid SSNs to be dropped from economic censuses processing.	
3. Spouse SSN	Source of identification for Minority-Owned and Women-Owned Businesses programs.	
4. Accounting Period	Identifies invalid tax years to be dropped from economic censuses processing.	
5. Name (first, second, and character counts)	Potential use for Standard Industrial Classification name-coding process.	
6. Address (street address, city, state abbreviation, ZIP Code, and character counts)	Source of geographic codes for nonemployers in the economic censuses.	
7. EIN	Provides identification/link to EI records, including SSEL.	
Additional Data Items (Applicable Form Types are Specified):		
1. Gross Profits--Cash (Form 1040, Schedule F)	Not used.	
2. Gross Profits--Accrual (Form 1040, Schedule F)		
3. Business Address Including ZIP Code (Form 1040, Schedule C)	Provides physical location for nonmails/nonemployers.	$975,000
4. PIA Code (Form 1040, Schedule C)	Source of industrial classification.	
5. Gross Receipts (Form 1040, Schedule C)	Combined to form net receipts, which is used as proxy data for 1987 Economic Censuses nonresponse or nonmail employers, as well as for all nonemployers. Also, used in edits of reported data.	
6. Returns and Allowances (Form 1040, Schedule C)		
7. End-of-Year Code (Form 1040, Schedule C)	Source of Business Division establishment counts and used in subject-matter division edits. Also, used in the current business surveys.	$45,000
8. Months Actively Operated (Form 1040, Schedule C)	Used in subject-matter division edits. Also, used in the current business surveys.	$70,000
9. Salaries and Wages (Form 1040, Schedule C)	As proxy data in the Survey of Minority-Owned Business Enterprises. Also used in nonemployer edits.	
10. Cost of Labor (Form 1040, Schedule C)		$150,000

10/   The costs of . . .

Data Items	Data Item Uses	Special Keyed Items 1/
Data Items Common to All Types Forms (1120, 1120-A, 1120F, 1120S, 1065, 990, 990C, 990PF and 990T):		
1. EIN	Provides identification/link to EI records, including SSEL.	
2. Accounting Period	Identifies invalid tax years to be dropped from economic censuses processing.	
3. Form Code	Distinguishes form types (e.g., 1120 and 1120S).	
Additional Data Items (Applicable Form Types are Specified):		
1. PIA Code (Forms 1120, 1120-A, 1120S, 1120F, and 1065)	Source of industrial classification.	
2. Net Receipts (Forms 1120, 1120-A, 1120S, 1120F, and 1065)	Provides proxy receipts data for the 1987 Economic Censuses nonmail or nonresponse employers, as well as for all nonemployers.	
3. Gross Royalties (Forms 1120 and 1120-A)	For selected kinds of businesses, royalties included as receipts.	
4. End-of-Year Code (Forms 1120S and 1065)	Basis of Business Division establishment counts for nonmails and used in subject-matter edits. Also, used in the current business surveys.	$10,000
5. Months Actively Operated (Forms 1120S and 1065)	Used in subject-matter division edits. Also, used in the current business surveys.	$15,000
6. Last Name, SSN, and the Number of Shares of Stock Owned by Each Shareholder (up to ten)(Form 1120S)	Determines race, sex, and ethnicity for Minority-Owned and Women-Owned Businesses programs.	$185,000
7. Net Income (Loss) From Royalties (Form 1065)	For selected kinds of businesses, royalties included as receipts.	
8. Net Farm Profit (Form 1065)	Not used.	
9. Last Name and SSN of Each Partner (up to ten) (Form 1065)	Determines race, sex, and ethnicity for Minority-Owned and Women-Owned Businesses programs.	$360,000
10. Gross Dues and Assessments (Form 990)	Provides proxy data for 1987 Census of Service Industries nonresponse or nonmail employers.	
11. Gross Contributions, Gifts, and Grants (Form 990)	Provides proxy data for 1987 Census of Service Industries nonresponse or nonmail employers.	
12. Total Gross Receipts (Form 990)	Provides proxy receipts data for 1987 Census of Service Industries nonresponse nonmail cases.	
13. Total Receipts (Form 990PF)	Provides proxy receipts data for 1987 Census of Service Industries nonresponse nonmail cases.	

Table F--QFR Data Items and Their Uses
Estimated Procurement Cost: $15,000
Approximate Number of Records Annually: 250,000

Data Items	Data Item Uses	Special Keyed Items 12/
1. EIN	Provides identification/link to EI records.	
2. Name (first, second, third, and fourth names)	Provides name for QFR mailing purposes.	
3. Address (street address, city, state abbreviation, and ZIP Code)	Provides address for QFR mailing purposes.	
4. Consolidated Return Indicator	Indicates a consolidated return was filed.	
5. Net Receipts	Determines sample codes for QFR.	
6. Net Income or Loss	Provides a measure of size.	
7. Total Assets Code	Indicates asset ranges. Used to determine sample codes.	
8. Sample Code	Identifies the QFR sampling rate. Used in the sample selection from the BMF.	
9. Final Return Indicator	Indicates a final return was filed.	
10. PIA Code	Determines major industry groups for QFR.	
11. Annual Accounting Period	Determines mail codes for QFR.	
12. Partial Year Indicator	Indicates a part-year return was filed.	
13. District Office Code	Determines service center distributions for quality control purposes.	

12/ The QFR sample file contains no special keyed items.

Table G--Form 851 Affiliations Schedules File Data Items and Their Uses
Estimated Procurement Cost: 13/
Approximate Number of Records: 14/

Data Items	Data Item Uses	Special Keyed Items 13/
1. EIN of Parent Corporation	Used to link subsidiary corporations with parent corporations on the SSEL for our Enterprise Statistics program.	
2. EIN of Each Subsidiary	Source of subsidiary identification.	

13/ We cannot accurately project the cost at this time.

14/ We cannot accurately project the number of records at this time.

Table A--SSA Birth File Data Items and Their Uses
Estimated Procurement Cost: $120,000
Approximate Number of Records Annually: 800,000

Data Items	Data Item Uses	Special Keyed Items 1/
1. Employer Identification Number	Provides identification/link to EI records, including SSEL.	
2. Birth Reason Code	Indicates reason for business activation. Possible use in business birth/death analysis.	
3. SSA Class Code	Identifies corporations for the QFR; determines legal form of organization code for the SSEL when other sources are not available.	
4. SSA Birth Date	Indicates date of business activation.	
5. SSA State and County Code	Provides geographic codes for use in current business surveys.	
6. SSA SIC Code	Supplies the SSEL with SIC codes for new single-establishment firms.	
7. Estimated Employment	Provides a measure of employment size for current business surveys when 941 payroll is not available.	
8. Name and Second Name } 9. Address	Provides name and address for current business surveys when no matching Business Master File record exists.	

1/ The cost of assigning and keying the birth reason codes, the SSA state and county codes, and the SSA SIC codes is shared by the Census Bureau and the SSA.

Table B--Form 1040 Self-Employment (SE) File Data Items and Their Uses
Estimated Procurement Cost: $10,000
Approximate Number of Records: 8 Million

Data Items	Data Item Uses	Special Keyed Items 2/
1. SSNs (primary and spouse)	Source of identification.	
2. Nonfarm Net Earnings (primary and spouse)		
3. Farm Net Earnings (primary and spouse)	Provides a measure of size for matching to sole proprietorships that file a Form 1040 joint return. Used in the Minority-Owned and Women-Owned Businesses programs.	
4. SE Income (primary and spouse)		

2/ The Form 1040 SE File will contain no special keyed data items.

Table C--Race and Sex Coding File
Estimated Procurement Cost: $9,000
Approximate Number of Records: 20 Million

Data Items	Data Item Uses	Special Keyed Items 3/
1. SSN	Source of identification. Used to link race and sex coded SSA records to our file of businesses.	
2. Sex Code	Provides sex type for Survey of Women-Owned Businesses.	
3. Race Code	Provides race type for Survey of Minority-Owned Business Enterprises.	

3/ The Race and Sex Coding File will contain no special keyed data items.

APPROXIMATE PROCUREMENT COST OF THE
1987 BUSINESS INCOME TAX FILES
FROM THE INTERNAL REVENUE SERVICE (IRS)

IRS Function	Approximate Cost
Total Procurement Cost	$1,030,000
I. Basic procurement costs associated with disclosure review, IRS programming, IRS coordination, and file	$150,000
II. Additional keying costs to obtain data items not keyed as part of the normal IRS revenue processing	$880,000
• End-of-year code	$60,000
• Months actively operated	$90,000
• Cost of labor	$75,000
• Wages	$75,000
• 941 indicator	$50,000
• Last name, social security number, and number of shares of stock for stockholders (from Form 1120S); last name and social security number for partners (from Form 1065)	$525,000
• Gross royalties (special keying required for Form 1120S only)	$5,000

CENSUS BUREAU USES OF ADMINISTRATIVE
DATA FOR SERVICE STATISTICS

Howard N. Hamilton, Carl A. Konschnik,
Paul S. Hanczaryk and Brian V. Greenberg

United States Bureau of the Census
Washington, D.C. 20233

EXECUTIVE SUMMARY

The Bureau of the Census is engaged in the largest expansion of coverage for
the service industries in the last 50 years. In particular, we will increase
coverage into the areas of communication, transportation, finance, insurance,
and real estate. With the expansion Census Bureau economic programs will be
representing about 98 percent of the Gross National Product (GNP), up from 74
percent currently. Coverage for the entire expansion will take effect for the
1992 censuses. As we expand coverage into these and related areas, the role
of administrative record data will increase.

The Census Bureau makes extensive use of administrative records in a wide
range of programs. These records are key components of the economic censuses,
especially for the retail and service industries. We use administrative
records to assemble mailing lists and obtain industrial classification for
businesses included in economic censuses and related surveys. Through the use
of administrative data we are able to publish a variety of statistics at a
comparatively low cost and with much less burden on responding business firms.

Information about small single-establishment employer firms with payroll below
specified cutoffs is obtained from administrative records. Although the
payroll cutoffs vary by kind of business, small employers generally include
firms with 1 to 4 employees. Data on receipts, payroll, and employment for
these firms are derived from administrative records, except for a 10-percent
sample of small employer firms which are mailed. For the 1987 Census of
Retail Trade, small below-cutoff employers contributed about $168 billion or
11 percent of the annual sales volume and about 682,000 or 28 percent of the
number of retail establishments. For the 1987 Census of Service Industries,
small employers contributed about $126 billion or 15 percent of the annual
receipts volume and about 936,000 or 15 percent of the number of service
establishments.

Information and data values for nonemployers are also obtained from
administrative records. The economic censuses of retail trade and of service
industries each produce publications of statistics for nonemployers and data
in these publications are based primarily on administrative records. For the
1987 Census of Retail Trade, nonemployers contributed $47 billion or about 3
percent of the annual sales volume, and about 916,000 or 38 percent of the
number of retail establishments. For the 1987 Census of Service Industries,
nonemployers contributed $96 billion in receipts or 11 percent of receipts and
about 4.6 million establishments or about 74 percent of the number of service
establishments.

At present, annual and monthly estimates for nonemployers in retail and services are derived from an area sample. Retail stores are generally located in areas where people shop and are typically very noticeable, whereas service establishments are often not located in a central commercial district. In particular, they may easily go undetected in the area sample canvassing. We anticipate a marked improvement in estimates for service businesses through an expanded use of administrative records in our annual survey programs with an ultimate goal of replacing the area sample with administrative records based data.

We anticipate more benefits from planned expanded use of administrative data in the service industries than in most other sectors of the economy. There are a higher proportion of nonemployers and small employers in the service industries than in other trade areas and these firms contribute a greater proportion to estimates of receipts. For these reasons we plan to increase our efforts to improve the quality and utility of the administrative records data wherever possible.