The Development and Use of Industry Data by the Social Security Administration

by Linda M. Dill*

Over the past few years the Social Security Bulletin has published a series of technical articles that describe various Social Security Administration (SSA) data files. This article provides an overview of SSA’s industry-related data files and statistical systems from both a current and a historical perspective. The author begins by explaining how SSA first collected business data from employers (starting in 1937) as a by-product of the requirement that employers report employee wages for benefit computation purposes. She describes the administrative methods by which the data are collected, SSA’s coordination of its activities with other agencies, the data collection forms used, the scheme by which the data are coded, and the employer files into which the data are classified. In her closing, the author provides examples of the various uses of the industry data and the ways that these data relate to SSA’s statistical program needs and to those of other agencies as well.

*Division of Statistical Operations and Services, Office of Research and Statistics, Social Security Administration. The author wishes to thank Barry Dyc for his substantial contributions to this article.

From the beginning of the Social Security program, it was recognized that industry statistics derived from administrative records used to process Old-Age, Survivors, and Disability Insurance (OASDI) employer registration and wage reporting would be of great value not only to the Social Security Administration (SSA), but to other government and private organizations as well. SSA has, for example, used industry data to evaluate the effects of proposed legislation on various groups of workers, and the National Cancer Institute is using the data in its epidemiological research.

A large number of technical terms and abbreviations are used throughout this article; a list of abbreviations is provided on page 44 and a Glossary of Program Terms can be found on page 51.

SSA has engaged in the industrial classification of employers’ activities for statistical purposes since the agency enumerated the first 3 million employers covered under the Federal Insurance Contributions Act (FICA) in 1936-37. Although the Internal Revenue Service (IRS) took over the enumeration process in 1950, the industrial classification operation has remained with SSA.

SSA collects and maintains coded statistical information on the employer’s industrial activities and other characteristics (such as geographic location and type of organization) under the authority of Title VII, Section 702, of the Social Security Act. Section 702 imposes on the Secretary of the Department of Health and Human Services the duty to study and make recommendations concerning effective methods of providing economic security. This mission is delegated to the Commissioner of Social Security and performed by SSA’s Office of Research and Statistics (ORS). The study of the revenue base for social insurance programs includes analyses of employment patterns for various industries.

In 1943, SSA instituted the Establishment Reporting Plan (ERP), a voluntary program designed to facilitate the processing of wage reports for large multiunit employers and to provide for the collection and classification of by-product
statistics on various employer characteristics, including primary industry activity, by establishment or reporting unit. Generally, SSA defines a multiunit employer as one who operates more than one place of business in different geographic areas or engages in different industrial activities under the same Employer Identification Number (EIN). Under the ERP, a multiunit employer with 100 or more workers (50 or more prior to 1984) who met at least one of the following criteria was asked to use the plan: (1) 6 or more employees in a secondary State, (2) 10 percent of the employment or at least 50 employees in a secondary county or industry, or (3) 2 or more manufacturing establishments in the same geographic area. For three decades, many multiunit employers elected to use the ERP. However, the plan was discontinued in 1991 because of declining employer participation due to changes in employers’ wage reporting requirements (that is, annual wage reporting and magnetic media reporting). There was an overall decline of almost 81 percent in ERP participation from 1971 (13,944 employers) to 1986 (2,675 employers). The discontinuance of ERP has resulted in SSA’s classifying all entities as single units, regardless of the number of establishments or reporting units, under one primary industry code with the EIN being both the identifier and the unit of classification.

On an annual basis, SSA collects industry data in its Employer Classification Files (ECF): the Multiunit Employer Code File (MUCF) is for ERP employers, and the Single Unit Employer Code File (SUCF) is for non-ERP employers. The Office of Research and Statistics uses these files to code jobs of workers in the Employee-Employer (EE-ER) files (both current and longitudinal) of its 1 percent Continuous Work History Sample (CWHS). These files are described later in this article. Beginning with the coding of the 1992 cycle of the CWHS, scheduled to take place in the spring of 1994, the SUCF will become SSA’s sole data base of coded industry information on employers (because of the discontinuance of ERP).
Individual industry codes for establishment or reporting units of each active EIN in the MUCF will be consolidated into the SUCF under either a single primary two-, three-, or four-digit industry coded record or an unclassified record.

Administrative Sources of Employer Industry Information

Currently, IRS's Form SS-4 (Application for Employer Identification Number) is the primary source document used by SSA to assign an initial industry classification for new employers (exhibit 1). With the exception of household employers and nonemploying entities with either "6-million series" or "7-million series" EIN's (such as trust funds, fiduciaries, and estates), SSA assigns industry codes to all SS-4's received from IRS for entities that are required under IRS regulations to file for an EIN. The number of annual SS-4's received from IRS is about 1.2 million. Of these, approximately 900,000 are processed with an industry code.

For various reasons, not all employer-completed SS-4's are received by SSA. Each year, to ensure that business birth records are established for all active (employing) EIN's in its employer files, SSA accretes records from the employer's annual wage reports (AWR) it processes for those EIN's for which an SS-4 was received from IRS. The initial unclassified industry code is generated for each of these records. This process provides SSA with EIN identifiers for missing SS-4's that were received from IRS and for those EIN's that were sent during SS-4 data entry.

An initial unclassified industry code is generated for each of these records. This process provides SSA with EIN identifiers for missing SS-4's that were received from IRS and for those EIN's that were sent during SS-4 data entry. An initial unclassified industry code is generated for each of these records. This process provides SSA with EIN identifiers for missing SS-4's that were received from IRS and for those EIN's that were sent during SS-4 data entry.

Because its resources for updating industry codes are limited, SSA has coordinated its coding activities with the Census Bureau. SSA provides Census with a list of active EIN's from the ECF, and, in return, Census provides SSA with industry and other statistical codes from the Standard Statistical Establishment List (SSEL). About two-thirds of the industry codes in the SSEL are from the quinquennial Economic Census reports. Other sources of industry data in the SSEL are (1) current business surveys of the Census Bureau, (2) business birth files from the Social Security Administration, (3) the Business Establishment List from the Bureau of...
Labor Statistics (BLS), and (4) Principal Business Activity Codes from the IRS.4 The most recent coordination effort, which took place in 1974, was primarily tied to Census and SSA’s coding needs resulting from the 1972 Standard Industrial Classification (SIC) revision; it did not involve any industry data from BLS or IRS sources. SSA has recently asked Census to participate in a code coordination activity that would enable SSA to obtain employer industry codes from Census’ SSEL. This list is currently updated with coded information from the 1987 Economic Census and, for more recent years, from the sources mentioned above. Since the SSEL is updated every 5 years with results from Economic Censuses, SSA intends to initiate additional coordination efforts. They will take place as coded results from future Economic Censuses become accessible in the SSEL.

Industry Coding Scheme

The initial industrial classification system used was developed internally at SSA (which was then known as the Social Security Board) and implemented in January 1937. Later, to attain comparability of statistical data with those of other Government agencies, SSA adopted the first Standard Industrial Classification in two phases—manufacturing in 1947 and nonmanufacturing in 1951. The SIC Manual was developed by an interagency technical committee (with major SSA representation) and published by the Bureau of the Budget (now OMB). This conversion of SIC codes to the SIC also opened the way for coordination of SIC coded records with those of other agencies (such as the Census) for the updating of SSA’s ECF. To date, SSA has participated fully in all revisions to the SIC. Through its Office of Research and Statistics, SSA is represented on the interagency Technical Committee on Industrial Classification (which meets under the auspices of the OMB). Revisions in the SIC are incorporated regularly into SSA files.

The current industry classification used by SSA is based on the 1987 SIC, which was implemented by the Agency in 1989. The scope of SSA’s industry classification covers all divisions of the SIC. The SIC enables SSA to classify primary business activities of EIN’s to a partial two-digit (major group) or three-digit (industry group) or a four-digit (industry) code according to the degree of detail in information provided on the source form. Partial codes are “zero filled” to the right to form a four-digit code. Unclassifieds are assigned “0000.” The zeros in the partial and unclassified codes are used to flag those employers who require correspondence for additional information.

The majority of the codes used by SSA are identical to those of the SIC. In some instances, however, either the full detail is not required for SSA’s statistical purposes or correspondence for additional information is not desired. In such cases SSA assigns foldback codes, which are consolidations of two or more SIC codes in related areas. Foldback codes are used only to classify some of the nonmanufacturing activities. For example, retail restaurants and bars, which are defined by the SIC under two separate four-digit categories (5812 and 5813) are classified by SSA under foldback code 5811. This assignment is made because these activities are commonly reported together on administrative sources without sufficient information to determine the primary activity. Another example is the industry

---

Exhibit 3

Exhibit 4
group, 151, for general contractors of buildings. The four-digit SIC classifications under this industry category require knowledge of the primary type of buildings being contracted. Since this level of detail is not normally supplied by the employer on the source form, a foldback code, 1511, is assigned. Currently, to expedite the industry coding provided by an employer, SSA uses the foldback code regardless of the level of detail supplied by the employer. Previously, with the exception of farming and public administration activities, the complete SIC code was applied in lieu of the foldback code when sufficient information was available on the source form. This practice required the classifier to go to an additional reference to obtain the SIC code.

SSA also uses special codes that deviate from the SIC: for example, code 9911 is for certain activities, such as building one’s own home, that do not fall within the scope of any SIC definition; code 4218 is used for SS-4 applications for highway use tax that do not give a primary activity; and code 6530 is used for some real estate activities that are coded from the employer’s trade name only. Prior to the 1987 SIC, special codes were also assigned by SSA to differentiate between the different types of agricultural crew leaders (for example, corporate and foreign in contrast with individuals and partnerships, those with equipment contrasted with those that supply workers only). These codes were used by ORS to publish agricultural statistics in earlier years of the program in order to evaluate the degree of compliance among crew leaders.

To prevent numerous unclassified industry codes, SSA applies internally established “forced coding” interpretations to certain industry descriptions to deduce a partial or complete industry code. For example, if an employer is engaged in the retail sale of a product that can be sold either from a stationary facility (such as a store), or through direct selling organizations, or by mail order, the industry activity is coded for sales from a store, unless it is otherwise designated by the employer. Also, except for the sale of automobiles, if the employer does not specify whether the products being sold retail are new or used, the industry code is assigned with the assumption that the products are new. Another example of forced coding is the activity of printing, which falls under two separate SIC manufacturing major groups (MG) for classification purposes. When not further described, this activity is force coded to the two-digit partial code that represents the MG for printing on paper or other materials, not including textiles or finished fabric articles.

SSA also force codes certain combinations of commonly associated activities to select one activity over another. If reported separately, these activities would be classified under different industry codes. For example, the combination of an automobile repair garage and towing service is coded as a repair garage based on the premise that the former is usually the primary source of receipts.

SSA’s forced coding policy decisions have been instituted over the years based on surveyed results of primary activities from employer correspondence replies.

---

**Exhibit 5**

**Social Security Employer Classification Update**

From: Office of Central Records Operations
P.O. Box 2115, Baltimore, Maryland 21225

In processing information obtained from Forms SS-4, Application for an Employer Identification Number, and from annual wage reports, it is necessary for the Social Security Administration to contact employers for whom we do not have adequate information to assign proper industrial or geographic classification. The data compiled from your response to this letter are used with wage and employment totals to develop statistical records of significant benefit to business, industry, government, and the public.

Please provide the information requested below, including a description of your primary business activity. Many businesses can be sufficiently described by a word or two, such as “restaurant” or “gas station.” If your business consists of two or more activities, then show an estimate of the percentage of gross income derived from each activity. The accurate classification of your business may depend upon details; therefore, please be as specific as necessary. As a guide, refer to the examples of business descriptions listed on the back of this page.

If your employer identification number and the address information shown above are not correct, please line through and correct the information.

Although a reply to this letter is voluntary, your cooperation is appreciated. When replying, please return the entire form in the enclosed business reply envelope.

Enclosure

---

1. Where does most of your business activity take place?
   - Town or City
   - County
   - State

2. Description of principal business activity (refer to examples on the back of this form)

3. Do you operate more than one place of business? [ ] Yes [ ] No

4. Number of Employees

Signature and Title

Telephone Number (include area code)

Date

Department of Health and Human Services
Social Security Administration

Form SS-4 (SFE-35, 1-91)

Social Security Bulletin • Vol. 55, No. 4 • Winter 1992
Manual was originally under ORS, but was transferred to OCR0 in 1964. Industry classification is a full-time operation staffed by approximately 30 people— including both classifiers and review technicians. OCRO gets coding policy guidelines, along with classification manuals, indices, and other industry coding aids, from ORS. OCRO’s coding procedure and software are developed by the Office of Pre-Claims Requirements (with ORS’s assistance).

Industry Coding Operations

SSA uses two primary publications to assign an industry code: the OMB’s Standard Industrial Classification Manual and SSA’s Industrial Classification Index (ICI). The ICI is divided into two alphabetized sections—manufacturing and nonmanufacturing—for purposes of classification. SSA also utilizes personal computer (PC) software for computer-assisted industry coding. The software provides a keyword index with items arranged alphabetically within each major SIC division for searching for and selecting an industry code. Fieldback codes are identified with asterisks in both the ICI and the PC keyword index; in the SIC Manual they appear within bracketed annotations, entered by SSA users. Index items used to apply forced coding are contained either in the ICI or on an unpublished list of precedent case code decisions. The list is used as a coding reference to interpret certain “not further described” industry descriptions often used by the employer, and it is also used to aid in resolving difficult cases.

SSA’s industry classification of employers is performed by the Office of Central Records Operations (OCRO). This activity (including the personnel) was originally under ORS, but was transferred to OCRO in 1964. Industry classification is a full-time operation staffed by approximately 30 people—including both classifiers and review technicians. OCRO gets coding policy guidelines, along with classification manuals, indices, and other industry coding aids, from ORS. OCRO’s coding procedure and software are developed by the Office of Pre-Claims Requirements (with ORS’s assistance).

Single Unit Employers

SSA receives SS-4 forms daily from IRS Service Centers. After the forms are microfilmed by SSA and given microfilm reference numbers, they are grouped by month of receipt for coding. The household and nonemploying “6-million series” EIN or “7-million series” EIN forms are excluded from the procedure.

Each new EIN is assigned one industry code which, in most cases, is based on information provided by the employer on the “principal activity” line of the SS-4. However, if additional information is required to assign an industry code or if the activity line is blank, other relevant information on the SS-4 is used to aid in determining the primary activity. Such information might include (1) true or trade name, (2) type of organization (for example, State/local government, nonprofit organization, National Guard, trust, Federal Government/military, church or church-controlled organization); (3) type of employees (nonagricultural, agricultural, household); (4) principal product manufactured and raw material used; and (5) identification of the groups or entities to whom most of the products or services are sold (business establishments, general public, other). If an industry code cannot be assigned to the MG (two-digit) level, employers with 100 or more employees are contacted by telephone for additional information; all others are assigned an unclassified industry code.

At present, about half of the SS-4 industry codes are assigned by manual methods and about half with computer assistance. With manual processing, the industry codes are recorded on each SS-4 under one operation and then, in a separate operation, the data are keyed for all cases with valid MG codes. Under the computer-assisted employer classification (CAEC) method, the computer software provides a one-step operation that aids the assignment of an industry code through the use of a keyword index search and at the same time provides both code validation and data entry capabilities. The CAEC program was partially implemented in 1989; full implementation awaits the development of a new main frame system that will provide on-line data entry for all SS-4 processing.

ERP Multunit Employers

During the period when the ERP was operational (1943–90), SSA contacted large employers who were identified as possible multunit employers through information received on the SS-4 and on annual wage reports and requested them to participate voluntarily in the ERP. Employers agreeing to participate were asked to complete and mail to SSA a form SSA-5019. In this way, the multunit employers could register those establishment or payroll reporting unit numbers under which they would be reporting employees on their annual wage reports and providing identifying information on each unit for industrial and other statistical classification purposes.

Form 5019’s were received from ERP employers and coded on a flow basis. Each of the ERP employer’s establishments or payroll groupings was manually assigned a separate industrial classification code based on information provided in the requested description of the business activity portion of the form. The description section consisted of (1) a check box in which to designate primary activity (that is, wholesale, retail, manufacturing, other), (2) a list in which to indicate, in order of importance, those products or services being produced under the primary activity, and (3) if there were multiple activities, a column in which to give percentages of gross dollar volume for each activity.

If an establishment’s primary activity was the provision of administrative or auxiliary services for other establishments of the same EIN rather than for the general public or for other business establishments, the employers were asked to specify the establishments being serviced. An industry code was then assigned based on the principal activity of the company unit(s) served. The subcode “1” was added if the establishment was to be designated either auxiliary or administrative. If there was not enough information provided on the form to
Employer Classification Files

The Social Security Administration currently houses industry codes for employers in two separate files—the SUCF and the MUCF. These files are maintained by the Office of Information Management for the Office of Research and Statistics.

Single Unit Employer Code File

The SUCF contains an EIN record for each form SS-4 that SSA receives. The file also includes records for EIN accretions from annual wage reports (form W-3’s) for those employers, except household, for which SSA has not received an SS-4. SSA assigns these accretions a computer-generated unclassified industry code that is retained in the file until an industry classification is received from another administrative source. In addition, the file contains (1) statistical codes for geographic location (State and county), type of organization (for example, individual, partnership, corporation, State/local government, Federal Government, nonprofit), SS-4 size, and reason for SS-4 application; (2) date and source of the codes; (3) W-3 size (number of W-2’s) for the five most recent years; and (4) FICA wages from the most recent W-3 year. Currently, the MUCF contains all active and inactive establishment or reporting unit number/EIN records processed by SSA, with the exception of inactive records that were purged from the file in 1974. Industry codes assigned to original or supplemental units in the file remained constant unless the employer notified SSA of a change in activity via a corrected or revised form 5019. Since SSA’s definitional structure of the unit of classification for multiunit (ERP) employers differs from that of other agencies, industry codes for units of ERP employers could not be updated through code coordinations with other agency records. Because of the discontinuance of ERP, the MUCF is no longer being updated.

One-Percent CWHS

Employee-Employer Files

The CWHS is a system of data files, maintained by ORS, that combines demographic, geographic, and longitudinal employment data for a 1-percent sample of Social Security numbers (SSN) issued. Industry data, along with other employer data from SSA’s Employer Classification File, are combined with individual worker data from annual wage reports and other administrative records to form several EE-ER data files that are part of the CWHS. When ECF codes are not available, other administrative sources are used to assign industry codes to workers in the CWHS. These sources include (1) type of employment codes from AWR’s for agricultural and household employers, (2) known EIN’s.
files also contain data on (1) other em-
ployers, (2) demographic characteristics of the workers, (3) wages, (4) insurance and benefit status, and (5) for 1983 and later, Medicare qualified government earnings (MQGE).

Auxiliary files prepared from the EE-ER files summarizing the data by major job industry categories at different levels include:

- **Assigned FICA file**, which contains one record for each wage earner who earned at least $1 in FICA-covered employment during the year. Industry codes are assigned based on the employer of the major job being the employer with the greatest FICA taxable wages;
- **Assigned FICA/Non-FICA file**, which contains the same as above, except that industry codes are assigned based on the employer of the major job being the employer with the greatest FICA taxable wages, total compensation, or MQGE;
- **Actual Industry File**, which contains one record for each two-digit industry in which the wage earner had employment during the reference year. The third and fourth digit of the industry code are assigned based on the employer of the major job within the same two-digit industry.

**Longitudinal EE-ER File**

The Social Security Administration’s 1-Percent Sample 1957 to Date Longitudinal Employee-Employer Data (LEED) file is a unique historical file in the Federal statistical community that contains industry data. The LEED file contains one record for each employer per year per SSN. It is constructed from all employee-employer combination records from each year’s 1-percent sample annual EE-ER file (except current year). Two versions of the file are maintained. The first, which is sorted by SSN, EIN, and year, groups together all industry-coded jobs by individual employee over the time span of the file. The second, which is sorted by year, SSN, and EIN, allows access to groupings of all employees and their industry-coded jobs reported under each year of the file.

**Uses of SSA’s Industry Data**

**Uses Within SSA**

The Social Security Administration uses industry data in the analysis and presentation of worker information for such purposes as evaluating the effects of proposed legislation, determining adequate coverage provisions, and disseminating basic statistics related to earnings and employment. Three examples of recent analytic uses of industry data from the 1-percent CWHS EE-ER files are (1) the development of industry data for aged and disabled Supplemental Security Income (SSI) recipients and for retired workers who are aged 65 or older, indicating the nature of their work experience; (2) the annual preparation of tabulations of workers insured for Disability Insurance benefits under the OASDI program, by major industry and age and by major industry and State, for analyzing State disability awards and program termination patterns; and (3) the targeting of the population of student workers, currently exempt from FICA taxes, for preparing estimates of additional revenues and of the number of student workers affected under various coverage test levels.

At the employer level, the SUCF recently was used to create a general purpose employer sample containing employer characteristic data (including industry) on active employers who filed wage reports with SSA. After linkage with other employer data from SSA’s wage reporting records, reconciliation records, and pertinent IRS tax records, the industry data could be used in assessing the impact of possible changes in the annual wage reporting process, responding to congressional and public inquiries, and assisting in program policy evaluations.

Although this article has focused almost exclusively on employer-based industry data, SSA has also used industry data from administrative resources other than the employer coding operations for dissemination of statistical information. For example, two-digit industry codes are assigned by each State’s Disability Determination Service, under Federal funding, on the Form SSA-831, Disability Determination and Transmittal. These codes are based on the industrial activity of the worker’s main predisability employer. SSA’s Office of Disability uses these codes to produce annual tabulations of new disability awards by diagnostic group and industry division. Also, primary industrial activity codes, self-coded by the taxpayer on IRS form 1040 (Schedules SE, C, and F) during revenue processing, are received by SSA on IRS self-employment tapes. The primary industrial activity codes are converted to SSA industry codes for self-employed persons whose Social Security numbers fall within the 1-percent sample for the CWHS and whose wages have been posted by SSA. These codes, along with other information on the self-employed,
Uses Outside SSA

Currently, the Census Bureau is the major outside user of SSA industry codes. Under an annual reimbursable agreement with SSA, Census receives, on a flow basis, tapes of coded business birth information from the SS-4 coding operation. Census uses SSA-assigned industry codes to supply the Standard Statistical Establishment List with SIC codes for new single-establishment firms.

Other recent outside uses include epidemiological research. In one study, the health affects of asbestos in the ship building industry during World War II were investigated. The National Cancer Institute (NCI) provided SSA with the Social Security numbers of persons who died from mesothelioma; SSA then supplied NCI with the work histories of these persons from the Employer Classification Files. In another study that also involved the NCI, SSA furnished industry data from the LEED file to produce tabulations showing the relationship between a deceased worker's years of exposure in industries commonly associated with excess mortality to the medical cause of death.

The longitudinal nature of the industry data in SSA's EE-ER files makes them a valuable resource of information for other potential uses. For example, the Health Care Financing Administration is considering using the industry codes assigned to workers in the longitudinal EE-ER files to analyze the relationship of Medicare utilization to workers from different industries.

Prior to the 1976 Tax Reform Act, the government and the private sector used the industry data extensively in the CWHS EE-ER files for such activities as (1) analysis of employment and migration patterns, (2) demographic, economic, and epidemiological research and studies, and (3) development of sampling frames for employer surveys. Whenever SSA released these files to outside users, all individual identifiers (such as Social Security number and name) were omitted to preserve confidentiality. With the passage of the Tax Reform Act, major new IRS confidentiality restrictions were imposed on the release of employer/employee information, especially microdata, from tax returns (forms W-2, W-3, and SS-4). SSA uses these data in its EE-ER files. These restrictions, however, do not prohibit SSA from releasing its statistical tabulations and summaries provided that they do not have identifiers.

Notes

5 Bureau of the Census, Department of Commerce, personal communication.
9 Industrial Classification Index, Office of Research and Statistics, Social Security Administration, 1988.
11 Industrial Classification Index, Office of Research and Statistics, Social Security Administration, 1988.
14 "To Date" here refers to the most recent year for which data are processed in the LEED file—usually 3 years behind the current date.

Glossary of Program Terms

Annual Wage Reports (AWR)

Prior to January 1978, employers reported FICA-covered earnings to SSA on a quarterly basis. Beginning with tax year 1978, employers were required to report annually using form W-2's. These forms are submitted with a transmittal form, W-3.

Continuous Work History Sample (CWHS)

The CWHS is a set of longitudinal data files containing information on a 1-percent sample of workers. These data files were developed to fulfill the need for statistics to be used in planning and operating the Old-Age Insurance system established by the Social Security Act of 1935. The contents of the files evolved
from information on employment, earnings, and benefits collected as part of the administration of Social Security programs. Today, the CWHS is still a major source of Social Security program statistics and workforce data. The CWHS data files are regularly used by SSA staff in making revenue estimates, evaluating legislative proposals, and responding to informational inquiries.

**Employer Identification Number (EIN)**

Each employer subject to the Federal Insurance Contributions Act is required to apply for an EIN using the Application for Employer Identification Number (Form SS-4). SSA codes information from the forms in order to classify jobs by industry and geographic location.

**ERP code file**

See Multiunit Code File.

**Establishment Reporting Plan (ERP)**

A voluntary program under which employers group employees by establishment or reporting unit within the annual wage report (AWR). Employers with more than one place of business and a total of 100 employees or more, of whom at least 6 are in a separate location, are asked to use this plan. The employers provide SSA with a list giving the location, industrial activity, and approximate number of employees for each establishment. The list identifies each group with a preassigned establishment number. Thus SSA is able to classify employees more precisely by geographic location and industrial type.

**Geographic code**

Five-digit numerical code, based on SSA’s Employer-Worker Classification Geographic Code Manual, assigned by SSA to employers for the purpose of geographically classifying jobs of workers in the CWHS. The first two digits of this code identify the State and the last three digits identify the county within that State.

**Industrial code**

Four-digit numerical code based on the 1987 Standard Industrial Classification Manual, assigned by SSA to employers for the purpose of classifying by industry the jobs of workers in the CWHS. The first two digits of this code identify the major group, the third digit identifies the industry group, and the last digit identifies the industry.

**Longitudinal Employee-Employer Data (LEED) file**

This file is assembled from the annual 1-percent sample of employee-employer (EE-ER) records. The original records from the various annual files have been resequenced so that all records associated with an employee, over the time span of the file, appear together. The basic data elements in this file include: (1) personal characteristics—year of birth, sex, and race; (2) wages—annual taxable, quarterly taxable (prior to 1978), total estimated wages (prior to 1978), and total compensations (1978 and later); (3) employer—State and county, industry, and coverage group; (4) insurance status; (5) benefit status; and (6) Medicare qualified government earnings (1983 and later).

**Major Group**

This code is comprised of the first two digits of the industrial code.

**Medicare qualified government earnings (MQGE)**

The earnings of government workers whose employment is not covered by Social Security but whose earnings are taxed for Medicare coverage.

**Multiunit Code File (MUCF)**

One record for each establishment (reporting unit) of all participating ERP employers is contained in this file. The record contains similar information at the unit level to that maintained in the Single Unit Code File (SUCF). Both MUCF and SUCF code files contain counts of the number of W-2’s reported for each of its five most current years for each EIN establishment. These data provide a means for measuring employer size and identifying active and inactive employers.

**Multiunit employers**

These employers have several physical locations and may be involved in diversified industrial activities.

**Non-ERP code file**

See Single Unit Code File.

**Single Unit Code File (SUCF)**

This file contains one record for each employer identified from a form SS-4 that is received and coded by SSA or that is accepted during annual wage processing when no record of an SS-4 receipt is in the existing SUCF. Excluded from the SUCF are household employers, nonemploying “6-million series” and “7-million series” EIN’s, and inactive employers whose records were purged from the files in the early 1970’s. Each SUCF record contains the EIN, industrial code, geographic code, business birth codes for coverage or type of organization, size, reason for application, and date and source of the information.

**Single unit employers**

These employers have one physical location and are involved in one primary industrial activity.

**Standard Industrial Classification (SIC)**

The industrial classification used for coding both multiunit and single unit employer forms in SSA is based primarily on the 1987 Standard Industrial Classification Manual.
Available Reprints From the Technical Article Series

Requests for copies of Social Security Bulletin Technical Articles should be addressed to the Social Security Administration, Office of Research and Statistics, Room 209, Van Ness Center, 4301 Connecticut Ave., NW., Washington, DC 20008; orders may also be placed by calling (202) 282-7138.

Development of Diagnostic Data in the 10-Percent Sample of Disabled SSI Recipients

The Social Security Administration has created a 10-percent sample data base of blind and disabled recipients from the Supplemental Security Income (SSI) program. Codes showing the primary medical diagnosis were obtained for the sample by matching several files, and by imputing codes to sample cases where no diagnostic codes were found. The data base is updated each year by repeating this matching operation and by bringing forward the diagnostic codes from the previous year’s file. This article describes the sources of diagnostic information in the administrative record system, the methodology used in the development of the 10-percent disability data base, and the technique chosen to compensate for missing values.

The Decline in Establishment Reporting: Impact on CWHS Industrial and Geographic Data

This article discusses the impact of the decline of employers’ participation in the voluntary Establishment Reporting Plan (ERP) program on the industrial and geographic data in the Continuous Work History Sample (CWHS), SSA’s largest file of employment and earnings records. Under the ERP, employers with multiunit businesses supply information that allows each unit to be classified under a separate primary geographic and industrial activity. In the mid-1970’s, participation in the ERP began to decline, and the decline has accelerated since 1978 when SSA changed to annual wage reporting. The CWHS area most affected involves geographic data; the impact on industrial classification has been less severe and primarily involves employers in the manufacturing sector.

The Social Security Administration’s Continuous Work History Sample

The Continuous Work History Sample (CWHS) is the oldest major longitudinal sample data source in the Federal statistical system. It was developed to fulfill the need for statistics to be used in planning and operating the Old-Age Insurance system established by the Social Security Act of 1935. This article discusses the current structure of the CWHS data system and files; it also describes some uses of the sample data and a number of issues that will affect the future of the system.

The Monthly OASDI One-Percent Sample File

This article briefly describes the development of SSA’s Master Beneficiary Record (MBR) and documents the contents, technical features, and uses of the 1-percent sample file that is developed from it. This sample file is designed to enhance the production of descriptive statistics, simulations of the beneficiary population, and other statistical analysis and research projects. A 1-percent sample file has been generated for each month since January 1985.