

STATISTICS OF INCOME FOR INDIVIDUALS: A HISTORICAL PERSPECTIVE

Jack Blacksin and Ray Plowden, Internal Revenue Service

Statistical aggregates based on information reported on Federal tax returns have been made available to the public since 1916. These data, in the form of published annual Statistics of Income (SOI) reports [6], have provided users with benchmark statistics on the distribution of income and tax for use in assessing the operations of the internal revenue laws and in estimating the revenue impact of proposed new tax legislation.

The present paper traces the history of the SOI program for individual income tax returns. We first summarize the evolution of the statistical series during the past 65 years, including a description of the tax law changes which took place during this period. Next we describe the types of data published and examine a number of economic and tax trends. There is then a discussion of the present procedures used to produce the data and planned procedural changes. The present and future outputs from the program are also indicated. Finally, appendix material is provided on SOI methodology. [7]

HISTORY AND BACKGROUND

The basic function of the IRS Statistics of Income series of reports, which is to provide "statistics reasonably available with respect to the operations of the Internal Revenue laws," is closely related to the history of taxation. Therefore a brief historical summary of the major developments in individual income taxation [2-4], which follows, will serve as an indication of the types of information the SOI report has been providing over the years.

The Sixteenth Amendment to the Constitution, enacted in 1913, provided that "Congress shall have power to lay and collect tax on income, from whatever sources derived, without apportionment among the several States, and without regard to any Census or enumeration". The individual income tax, enacted shortly afterward, applied to wages and salaries, interest, dividends, rents, entrepreneurial income and capital gains; the law also allowed deductions for personal interest and tax payments as well as for farmers' expenses; and exempted from all tax Federal, State and local Government bond interest, salaries of State and local Government employees, and dividends (from normal tax, but not from surtax). The tax law also provided for an exemption of \$3,000 for single persons and \$4,000 for married couples. Tax rates consisted of a 1 percent normal tax on taxable income plus a surtax ranging from 1 percent to 6 percent on net income over \$20,000.

Significant changes made between 1913 and 1965, those which mainly established the basic system of tax assessment, have been the allowance of a

credit for dependents and a deduction for charitable contributions (1917); elimination of collection at the source (1916) and reenactment of income tax withholding on wages and salaries only (1943); adoption of preferential tax rates on long-term capital gains (1921); elimination of exemption from taxation the salaries of State and local Government employees and discontinuation of the sale of tax-exempt Federal bonds (1941); adoption of the standard deduction (1944); enactment of "income splitting" for married couples (1948); introduction of self-employed pension deduction (1963); and introduction of income averaging tax method and minimum standard deduction (1964).

Major tax law changes made since 1965, reflecting more of a concern with social rather than revenue objectives, included establishment of a system of graduated tax withholding for salaries and wages (1966); allowance of partial deductibility of premiums for medical care insurance without being limited by adjusted gross income (1967); replacement of the minimum standard deduction by a low-income allowance, liberalization of tax return filing requirements for low income persons, and imposition of a minimum tax on specified income subject to preferential treatment (1970); introduction of a maximum tax on earned income (1971); introduction of the work incentive credit and the Presidential Campaign Fund checkoff (1972); revision to pension and employee benefit rules (1974); establishment of the earned income credit and the allowance of an adjustment for an Individual Retirement Account (1975); institution of a child care credit, general tax credit and credit for the elderly--previously the retirement income credit (1976); establishment of the new jobs credit, currently called the targeted jobs credit (1977); institution of residential and business energy investment credits (1978); and establishment of the advance earned income credit and the taxation of unemployment compensation (1979).

At the early stage of this statistical series, tabulations were few in number and relatively uncomplicated. With the passage of time, and as the tax laws became more complex, the statistical program has reflected demands by users for more data with an increasing degree of tabular complexity.

For the earliest SOI report of 1916 the information was summarized into just seven basic tables and 137 pages. Topics covered were income by source, occupation, tax by source, sex and marital status; size classifications were provided for income, and statistics were tabulated at the National, State and Territorial levels.

Some 30 years later, the SOI report for 1946 had increased to fourteen basic tables, largely through the introduction of cross-tabulations. New data added along the way covered such items as more detailed sources of income (including gain from sales of capital assets), standard and itemized deductions (including types of itemized deductions), number of exemptions and tax payments. New classifiers employed were filing status, size of specific income sources and net income classes.

By 1979, the SOI report had shown a still greater increase in the number of basic tables. Added detail, for instance, is being shown for counts of number of returns filed, sources of income, marital status and for nontaxable returns. There is more information for types of dependents, types of tax computation and for several types of tax credits. Examples of new items are statutory adjustments and minimum tax. New classifiers include age 65 and over, marginal tax rates and alternative income concepts. Data from tax return schedules are also being tabulated, such as income averaging, tax preferences, residential energy investments and energy credit.

ECONOMIC AND TAX TRENDS

One reason for the increasing usefulness of SOI as a statistical series has been the increase in the coverage of the income tax. For most of the first 25 years in which the income tax was in effect, the number of returns ranged between three million and seven million. The number, however, doubled from 7 to 14 million between 1939 and 1940 when stricter filing requirements were enacted. By 1950 it had passed 50 million; since then the number has increased by an average of more than ten million every decade and, by 1979, reached over 92 million returns.

Year	Returns (Thousands)	Year	Returns (Thousands)
1916	437	1950	53,060
1920	7,260	1960	61,028
1930	3,852	1970	74,280
1940	14,711	1979	92,694

In 1916 perhaps one out of ten persons in the population was represented on tax returns. By 1940 this number had risen to one out of three. Currently about 95 percent of the population is represented on tax returns as either a taxpayer, spouse, child, or other dependent.

The amount and nature of income from taxable sources has grown greatly. For example, "adjusted gross income" (consisting of all taxable sources, positive amounts less negative amounts and allowable adjustments, reported on tax returns), increased from \$150 billion in 1974 to \$1.5 trillion in 1979, about 10 times. The average income per return in constant 1972 dollars increased from \$6118 in 1947 and reached \$8957 in 1979.

The distribution of returns by size of adjusted gross income has been changing greatly during the past three decades (see Figure A). The number of returns with income under \$5,000, previously the largest single group, declined from 51 million in 1947 to 21 million in 1979. By way of contrast, the number with income \$5,000 under \$10,000 increased during this period from 3 million to 20 million. The largest growth was registered for the number of returns filed with current dollar incomes of \$10,000 under \$50,000 which now constitute over half of all returns filed. Some 2.3 million returns (2.5 percent of all returns filed) reflected income of \$50,000 or more for 1979.

Historically, the major sources of income, in order of importance, are salaries and wages, interest, income from sole proprietorships and partnerships, dividends, and net capital gain. For most years these five sources of income accounted for over 97 percent of total adjusted gross income. Salaries and wages, the largest single source, increased both in amount and as a proportion of income, representing since 1930 more than 60 percent and, since 1952, more than 80 percent of adjusted gross income. Interest income, once the smallest of the major income sources, increased the most rapidly, having exceeded dividends since 1967 and, for the first time in 1979, replacing income from sole proprietorships and partnerships, as the second largest income source. Income from sole proprietorships and partnerships, undoubtedly reflecting a shift to the corporate form of ownership, increased at a less rapid rate than most of the other major sources, and its proportionate share of the total declined sharply, from 42 percent in 1916 to under 5 percent in 1979. Dividends and net capital gain, the smallest of the major sources of income have, together, for the past forty years, accounted for about 4 or 5 percent of total income.

Internal Revenue Service Regions reflect the variation in return filing patterns (paralleling the variation in population growth patterns) that exist for different areas of the country. As shown in Figure B and Table 1, the number of returns filed in the West and Southwest increased at a rate well above the U.S. average. Since 1950, the number of returns in each of these regions increased by more than 100 percent. However, the largest increase took place in the Southeastern states where the number of returns filed almost tripled. The North-Atlantic Region, which included New England, New York and Puerto Rico, indicated the lowest rate of increase, about 26 percent. In the Mid-Atlantic, Central and Midwest Regions, the growth in number of returns filed (42 to 50 percent) was somewhat lower than for the Nation as a whole, 76 percent. The only place that showed an absolute decline in the number of returns was the District of Columbia which dropped from 373,000 in 1950 to 316,000 in 1979 (i.e., by 16 percent).

FIGURE A.--GROWTH IN NUMBER OF RETURNS BY SIZE OF ADJUSTED GROSS INCOME, 1946-1979

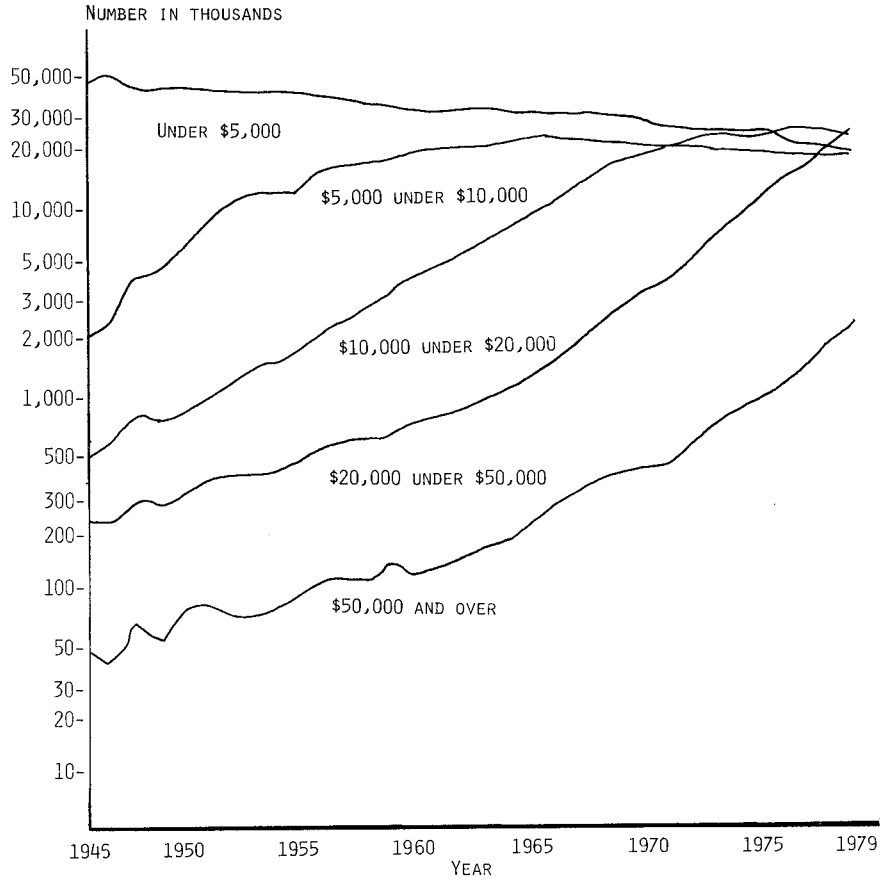
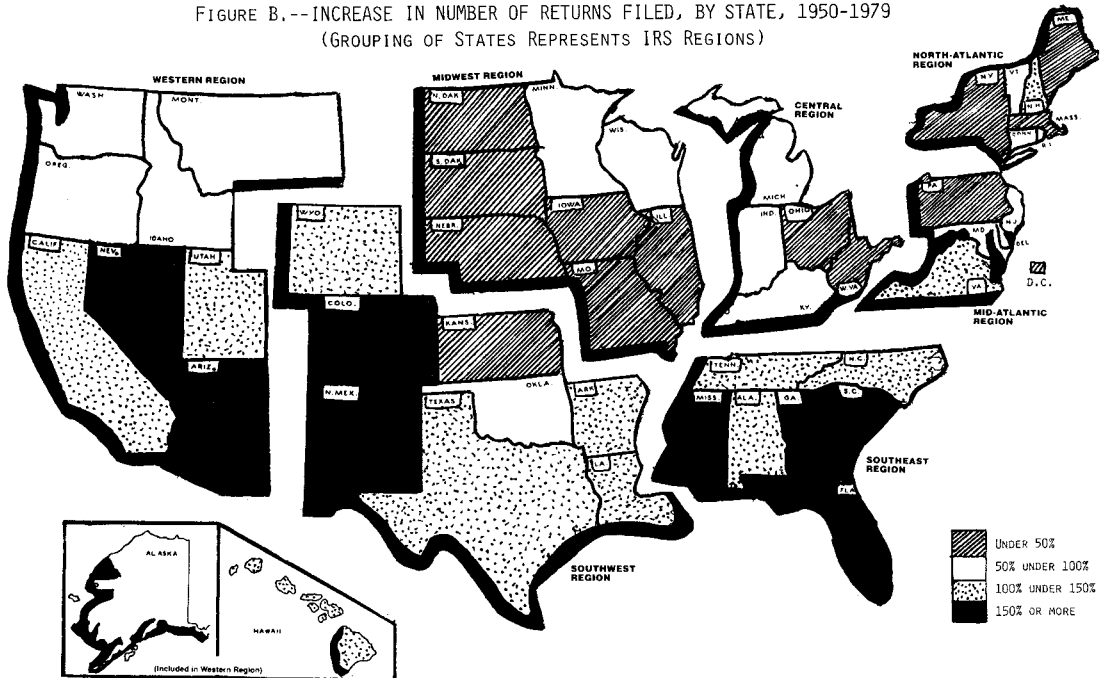


FIGURE B.--INCREASE IN NUMBER OF RETURNS FILED, BY STATE, 1950-1979
(GROUPING OF STATES REPRESENTS IRS REGIONS)



NOTE: The number of returns filed for Washington, DC, declined during this period by 16 percent.

Table 1.--Number of Individual Returns Filed by Internal Revenue Service Regions, 1916-1979

IRS regions	Number of returns (in thousands for selected years)								Index for 1979 (1950 = 100)
	1916	1920	1930	1940	1950	1960	1970	1979	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
United States..	437	7,260	3,708	14,711	52,665	61,025	74,278	92,694	176
North-Atlantic.....	140	1,753	1,054	3,445	9,911	10,529	11,728	12,513	126
Mid-Atlantic.....	80	1,300	677	2,634	8,689	9,322	10,913	12,545	144
Southeast.....	15	341	162	752	4,736	6,850	9,504	13,461	284
Central.....	46	1,117	471	2,280	8,323	8,952	10,532	12,484	150
Midwest.....	79	1,349	651	2,685	9,132	9,679	11,052	12,994	142
Southwest.....	29	627	258	1,054	5,248	6,632	8,443	12,044	229
Western.....	35	773	437	1,860	6,626	8,894	11,702	16,188	244

NOTE: States included in each IRS region are shown in figure B.

PRESENT PROCEDURES

Throughout the history of this statistical series, SOI data were based on stratified probability samples of unaudited individual income tax returns. The number of returns included in the sample has been over 1,000,000 (e.g., for 1948); for 1979, however there were about 200,000 returns selected, with smaller samples slated for later years.

During the first forty years of production of SOI, data were processed with electric tabulating machines (including manual key-entry equipment), with a coding or "edit sheet" employed as the standardized keypunch document.

The system used to produce SOI data became computerized in the 1950's, leading the rest of the Internal Revenue Service in automation by about ten years. In addition, computer designation of the sample was implemented in 1966. However, manual abstraction of data to an edit sheet and manual key-entry continued.

A further breakthrough in computerized operations for the individual SOI program came in 1974 with the introduction of a computer system which was patterned after a somewhat similar system used by the Canadian Department of National Revenue and Taxation to produce Taxation Statistics, their version of SOI [10]. This system permits utilization of an SOI file that combines both computerized data already available in the IRS Individual Master File (IMF), a tape record of data from the complete universe of individual income tax returns, with manually edited data available only from the tax return. Further enhancements were made to this system for Tax Year 1977 permitting consistency testing and error resolution of the statistical data within each of the ten IRS Service Centers where SOI basic processing is conducted and where returns are accessible for review during error resolution.

PLANNED PROCEDURAL CHANGES

The present processing system permits use of limited data already available on the IRS Master

Files, thereby practically eliminating the need for manual editing and key-entering Master File data for statistical purposes. However, data are still electronically printed onto an edit sheet, and additional data (some of which may be already in the Master File system) must still be manually edited.

Current plans now call for going a step further in data abstracting techniques. All data available in the Master File are now being reviewed with the aim of either foregoing the use of an edit sheet or creating a mini-edit sheet only for those items needed in the SOI program, but not part of the Master File system. This approach to processing would require more comprehensive consistency testing of all data at the service centers.

In the light of recent budgetary constraints, the Statistics Division has been forced to cut sample sizes. Steps are being taken, however, to maintain the reliability of the key statistical series. Among the methods being considered to preserve the reliability of estimates of year-to-year change is the greater use of longitudinal samples (already a key feature of the SOI designs for partnerships and corporations). Increased post-stratification to Master File totals will be explored as a way of preserving the quality of the overall cross-section statistics.

Other long-run changes being explored include doing more concurrent (or even on-line) consistency testing of the SOI records along with revenue processing, performing longitudinal consistency testing, and producing "print ready" tables from the computer.

Research is also being conducted on ways of speeding up production of SOI reports by closing out the file at an earlier date and then publishing the regular annual complete report based on this preliminary file [1]. Processing of the entire sample file would still be accomplished as in the past, but made available only on tape for use in special studies or as microdata records.

BASIC OUTPUTS

The SOI data are currently being provided in three separate documents: an unpublished Advance Data report, and published Preliminary [7] and Complete Reports [8]. In addition, using the same source documents, unpublished special tabulations are prepared such as Revenue Sharing Tax Liability, indications of Foreign Bank and Trust Accounts, line-by-line summaries of the frequency of entries on individual returns and attachments, and high-income tabulations. Special studies are also prepared for State governments, organizations such as the World Bank and for other Federal Government agencies for analysis [3]. Monthly and annual data for the Nation and by State are being provided based on the IMF. Further, the IMF is a source of Small Area Data (counties and SMSA's) provided for 1972 and 1974 [9] and is being considered for providing similar data for other years (1976, 1979, 1982, and 1984). Most of these studies were done at the request of specified users, many on a reimbursable basis.

Since 1966, SOI data have been provided annually in the form of a Tax Model, a microdata file containing unidentified records of returns selected for inclusion in the SOI sample. The Tax Model is retained in the National Archives and is also made available as a public use file [11].

In 1981, the IRS Statistics Division was gearing up for further changes in statistical output based on requirements registered by users. A major revision was contemplated in the use of Master File records to augment data, already available annually, on a more frequent basis, such as quarterly or monthly. In addition, a new set of statistics on occupations of taxpayers (possibly linked with age, race, and sex) based, in part, on Social Security Administration files, was being planned [5]. This study will be done partially on a reimbursable basis, for use in determining industrial mortality and morbidity rates. Other studies are anticipated, also on a reimbursable basis, such as the classification of more data by small geographic areas such as county and SMSA.

ACKNOWLEDGEMENTS

This paper is, in part, an update of material originally prepared by Helen Demond and included in her report, Historical Summary of Statistics of Income, 1916-1965, which the authors gratefully acknowledge. In addition, we wish to thank Ross Summers for reviewing the draft and providing helpful comments; Wendy Alvey and Beth Kiess for their help in presenting this paper at the meetings, and Clementine Brittain, Toni Jones-Lyles, Vicki Pazulski and Pat Piet for preparing the tabular material. Thanks are also due Ruth Wise who typed the tables and Cathy Robinson and Mary Haigler who typed the several drafts of this paper.

NOTES AND REFERENCES

- [1] Dumais, James and Raymond Shadid, "Individual Statistics of Income: Advancing the Closeout Date", 1981 Proceedings American Statistical Association, Section on Survey Research Methods.

- [2] Goode, Richard, The Individual Income Tax, Brookings Institution, 1964.
- [3] See, for example, Lerman, Allen H., High Income Tax Returns, 1975 and 1976, U.S. Treasury Department, Office of Tax Analysis, August 1978.
- [4] Pechman, Joseph A., Federal Tax Policy, Brookings Institution, 1966.
- [5] See especially Sailer, Peter and Harriet Orcutt and Philip Clark, "Coming Soon: Taxpayer Data Classified by Occupation", 1980 Proceedings American Statistical Association, Section on Survey Research Methods.
- [6] Statistics of Income, 1916-1979, Internal Revenue Service.
- [7] A summary of major tax law changes for individuals, types of data published in individual SOI reports, and detailed historical tabulations are available in Statistics of Income and Related Administrative Record Research, Internal Revenue Service, 1981, pp. 97-115.
- [8] Beginning in July 1981 the preliminary tables were published in the Statistics of Income Bulletin, Volume 1, Number 1, Internal Revenue Service, Summer, 1981.
- [9] Analytical applications of data included in the SOI Complete Report may be found in a report by Steuerle, Eugene and Michael Hartzmarck, Individual Income Taxation 1947-79, OTA Paper 48, April 1981.
- [10] Supplemental Report, Statistics of Income, Small Area Data, Individual Income Tax Returns, Internal Revenue Service, 1972 and 1974.
- [11] Taxation Statistics, Department of National Revenue-Taxation, Ottawa, Canada.
- [12] Use of the available files is described in a report by Wycarver, Roy A., The Treasury Individual Income Tax Simulation Model, Office of the Secretary of the Treasury, Office of Tax Analysis, Fall 1980.

APPENDIX ON METHODOLOGY

In recent years, the need for more statistical data from individual income tax returns has increased due to the passage of new tax legislation, the Revenue Sharing Program, growth in the number of Federally funded social and economic programs, and increasing utilization of microdata files for research in lieu of hard copy tabulations. Tax return records, in response to user needs, are presently utilized to produce a number of statistical reports and studies which are described below.

Basic SOI Program - The returns used in compiling the reports for the basic SOI program are from a sample of all 1040/1040A returns processed to the Master File System at the IRS National Computer Center. After selection of the sample, identifying information and some key data from the "Returns Transaction File" (part of the Master File system) are entered on a Statistics of Income tape known as the "Sample Receipts Tape (SRT)." The SRT tapes are sent to ten IRS Service Centers which processed the tax returns. Once at the Service Centers, these SRT tapes become part of a statistical processing system which combines both the data from the SRT

tapes with data that were manually edited and transcribed for statistical purposes. The merged record is then subjected to generalized consistency testing and error resolution. After the return records have been perfected, they become part of the "Composite Tape File" and these tapes are shipped to the IRS Data Center (in Detroit, Michigan) for further processing. At the Data Center, the return records are further perfected prior to table production.

The first two major tables produced are for an in-house Advance Data report. Prior to Tax Year 1979, the first published report, the "Preliminary Report" was prepared from this same early cut-off file. The preliminary report contained approximately twenty tables including the two basic tables produced for Advance Data. The "Complete Report" which is prepared using all returns obtained for the sample during the processing year, has consisted of about sixty tables, including final versions of those prepared for the Preliminary Report

Small Area Data Supplement - The geographic breakdown shown in the regular SOI reports is by States. The Small Area publication shows data by counties and Standard Metropolitan Statistical Areas for selected years. Also, unlike the regular SOI reports, this study has in the past been based on all returns in the IMF population.

In addition to the Master File data used in preparing this report, use was made of four Bureau of the Census files. One file contained county codes for all taxpayers while the other files contained either the "Geographic Indicator Check", the "Census Tally", or the "Standard Metropolitan Statistical Area (SMSA) Definition". The Geographic Indicator Check provides a list of geographic codes used by the Bureau of the Census as well as the correct spelling of the names of all counties and States. The Census Tally File shows the number of returns for each combination of ZIP Code and geographic code used on the tax return classified by ZIP Code. The SMSA Definition File contains a list of counties or minor civil divisions which are components of each SMSA as well as the name of each SMSA. These three small files were combined into a "geocode" file which contained all the codes and names necessary to produce the Small Area Data Report.

The hard copy of the Small Area report is not the only publication from the study. Typically a computer tape version of the summary data included in the tabulations is also made available. There is a plan to provide this file to the National Technical Information Service which will distribute tape file copies to interested users.

For future publications of this report, in an effort to save resources, a tentative decision has been made to compile Small Area statistics using a sample of the Individual Master File returns. For 1976 and for 1979, Small Area Data will be produced by taking a straight ten percent sample of the IMF.

Occupation Supplement - A new data base is in preparation which will contain not only the standard 1040/1040A tax return data, but also occupation of taxpayer (Standard Occupation Classification codes), industry of employer (Standard Industrial Classification codes), separate earned income data for husband and wife on joint returns, sex and possibly also age and race of the taxpayer. This data base, which is being financed partially by the National Cancer Institute, will be used to produce supplemental reports to the Statistics of Income series, and will also be available for research purposes.

At least for the first few years, the starting point for the occupation file will be the regular, individual Statistics of Income sample. Limited data from the Sole Proprietorship SOI file--industry codes for self-employed individuals, net business income for husbands and wives--will also be included. Form W-2 information will be obtained for the study, including salaries and wages of each taxpayer and employer identification numbers (EIN's). Using these EIN's, SSA will be able to provide industry codes for each taxpayer's employer.

Returns in the current SOI sample (and thus the occupation-coded file) contain five four-digit SSN ending digits (approximately 47,000 returns) common with the Continuous Work History Sample, and there is a plan to keep these returns in the future SOI samples as a panel. The National Cancer Institute would prefer a much larger panel (at least the one percent CWHS) for its studies. It is anticipated that by FY 1984 the complete one percent CWHS, drawn from the IMF transaction tapes at the National Computer Center, could form the basis of the occupation file. To accomplish this end, the taxpayer's entry in the occupation box for these returns will need to be keyed, either during or subsequent to revenue processing. These alternative coding schemes are presently being discussed within the IRS to determine the feasibility issues and cost effectiveness of each scheme.

A planned strategy is to automate the system of occupational coding of returns, to the extent possible, by creating a computerized dictionary of occupation titles used by taxpayers on their tax returns and the corresponding Standard Occupation Classification (SOC) codes. In some cases, the dictionary will contain multiple SOC codes for the same occupation title, with the choice of code based on the industry in which the taxpayer works. In order to verify this system of occupation coding, Social Security Administration requested that the Census Bureau interview a number of individuals as to their occupation, as part of the proposed Survey of Income and Program Participation. These individuals, identified by SSN digit endings, are included in the 1979 SOI occupation file. Census has been asked to make the results of the (roughly 800) interviews available to the Statistics Division for purposes of making direct case-by-case comparisons of the differing ways of obtaining occupation data.