
Statistics on Federal Taxation: The Statistics of Income Program of the IRS

Tom Petska, Internal Revenue Service *

This article describes the origins, customers, operations, and programs of the Internal Revenue Service's (IRS) Statistics of Income (SOI) function. In the first section, a brief introduction to the origins of the SOI program is provided. In the second section, SOI's customers and their statistical needs are described. In the third section, the operations and procedures of the SOI program are briefly summarized. Next, brief overviews of the SOI statistical programs are provided. Finally, some current limitations of SOI statistical systems are described, followed by discussions of current efforts to improve these systems.

■ Origins of the SOI Function

Under the U.S. self-assessment taxation system, taxpayers, whether individuals or businesses, report their financial affairs and calculate their tax liabilities on tax returns and related informational documents. Although this source of information is quite different from that of survey-oriented statistical organizations, the SOI program still has a similar mission -- to collect and process data so that they become meaningful information and to disseminate this information to its customers and users.

Statistical operations at IRS began about 80 years ago, with the passage of the sixteenth amendment to the U.S. Constitution (in 1913) and, later that year, the establishment of the first modern U.S. income tax law. Subsequently, the Revenue Act of 1916 required the annual publication of statistics, establishing a role for the SOI function. Despite many revi-

sions to the tax law, the original requirement of that Act continues today. Specifically, the current Internal Revenue Code states that "The Secretary (of the Treasury) shall prepare and publish, not less than annually, statistics reasonably available with respect to the operations of the internal revenue laws..." [1].

The costs of administering the Federal income tax system are substantial; the annual budget of the IRS for the current (1995) fiscal year is \$7.9 billion. Despite its mandate for producing statistics on taxation, the SOI program budget is only about \$30 million (0.4 percent of the IRS total). Thus, the SOI program represents a very small portion of IRS resources, and the SOI mission is clearly different from the overall IRS mission.

■ Customers of SOI Data

Since the mandate for the SOI program has a responsibility required of the Secretary of the Treasury, it is not surprising that SOI's primary customer is the Office of Tax Analysis (OTA) in the Office of the Secretary of the Treasury. Another primary customer is OTA's legislative branch counterpart, the Congressional Joint Committee on Taxation (JCT). OTA and JCT use SOI data as their primary source of information for analysis of the functioning of the tax system. In both agencies, microsimulation modeling is employed using SOI data as the primary database for policy analysis and tax revenue projections [2]. The SOI data are also sometimes matched with other data to build comprehensive databases that can be used in estimating the overall impact of tax law changes and their effects on tax collections.

Although the bulk of SOI's resources are focussed on the statistical needs of OTA and JCT, SOI has many more customers. The Department of Commerce's Bureau of Economic Analysis (BEA) is a significant user of SOI data for estimating components in the national income and product accounts. Corporation net income (less deficit) is used in the

* Tom Petska is Chief of the Coordination and Publications Staff, Statistics of Income, Internal Revenue Service (CP:R:S:P), P.O. Box 2608, Washington, DC 20013-2608. This paper was prepared for presentation at the Mid-Year Meeting of the American Tax Association, February 1994, in Atlanta, Georgia.

national accounts to benchmark corporate profits. Net income (less deficit) from proprietorships and partnerships is used to estimate unincorporated business income.

The Census Bureau, also in the Department of Commerce, is another significant data user; however, its needs (unlike those of OTA, JCT, and BEA) are primarily for IRS master file data (i.e., population data), rather than the lower volume but content-rich SOI samples. (This distinction is described later.)

Many other Federal agencies are also users of SOI data, including the Federal Reserve Board, the Environmental Protection Agency, and the Health Care Financing Administration. In each of these uses, IRS tax data are a rich (and sometimes only) source of information on the financial activities of the study populations. As with BEA and Census, however, these data needs are often fulfilled as a result of reimbursable contracts, since the resources of SOI are heavily committed to meeting the informational needs of OTA and JCT.

In addition, SOI data are used by a broad array of tax practitioners and researchers, State and local governments, the media, the public at large, as well as for internal operations in IRS. Figure A provides a summary of current requests for data from the public.

■ Statistical Operations

SOI statistical systems are not unlike those of other statistical organizations. Sampling is a major tool in study designs, and computers are an ubiquitous element. Data collection is a highly structured and disciplined process. Sample estimates are usually obtained by randomization-based weighting of selected cases. Aggregate tables are compiled and published, and public-use (microdata) files are made available after being "sanitized" to satisfy disclosure concerns. Research on methods, often driven by operating concerns, is conducted in ongoing attempts at improvement [3]. This section describes how these statistical activities are applied in SOI programs.

Figure A.--Statistics of Income Information Requests, by Type of Requester, 1994

| Inquirer | Telephone | Written request |
|--------------------------|-----------|-----------------|
| Total | 100.0% | 100.0% |
| Consultant/researcher | 18.2 | 17.2 |
| Accounting firm | 3.8 | 3.8 |
| Association | 7.3 | 11.2 |
| Law firm | 3.4 | 4.3 |
| Other private business | 8.8 | 27.4 |
| College | 7.0 | 7.9 |
| Public library | 0.6 | 0.7 |
| Private citizen | 7.4 | 10.2 |
| State/local Government | 6.8 | 9.0 |
| Internal Revenue Service | 15.9 | 1.1 |
| Congress | 5.5 | 1.7 |
| Other Federal Government | 7.9 | 1.4 |
| Foreign | 0.5 | 1.2 |
| Media | 5.3 | 2.5 |
| Student | 1.6 | 0.5 |

Note: Detail may not add to totals because of rounding.

Sample Design and Selection

U.S. tax returns are filed and administratively processed at one of ten IRS regional sites, called "service centers." Once processed, IRS compiles selected information from all returns into a computerized "master file" system, which is the informational backbone of the agency. SOI operations begin by sampling returns from the master file system; the master file offers a sampling frame that enables use of sophisticated and efficient sample designs.

Statistics compiled for the SOI studies are generally based on stratified probability samples of tax or information returns. As returns are processed into the master file system, they are assigned sampling strata, based on criteria such as income (or other measures of economic size), industry, or presence of supplemental forms and schedules.

Each taxpayer, whether an individual or a business, has a unique number -- the social security number (SSN) for individuals or the employer identifi-

cation number (EIN) for businesses. These unique taxpayer identification numbers (TIN's) are used as the seed for a pseudo-random number (a transform of the TIN) which, along with the sampling strata, determines whether a given return is to be selected into the sample (Harte, 1986).

The algorithm for generating the TIN transform stays the same from year-to-year. Consequently, a return is selected into the SOI sample if it falls into a stratum with the same or higher probability of selection. If it falls into a stratum with a lower selection probability, the likelihood of selection will correspond to the ratio of the second year to the first year's selection probabilities.

Data Capture Techniques

After sampling, data items based on the electronically-available information from the master file system are substantially augmented with additional items captured from the (largely) paper tax returns. Statistical abstracting can take as little as a few minutes for a simple return, to as long as several days for a large corporate return.

Until a few years ago, SOI conducted basic informational processing in a "batch-mode," involving several operational units at all ten IRS service centers. Within each center, employees in different functional units manually coded, abstracted, key-entered, and error-corrected data from tax and information returns. This fragmented system increased handling costs and denied "ownership" and was not conducive to maintaining high levels of efficiency and quality.

To improve the data processing operations, a network of minicomputers was built solely for SOI statistical processing. This system uses on-line "transaction processing," so that all data capture operations are completed in one pass. In addition to reducing handling costs and removing overlapping responsibilities, accountability has improved, because one person is now responsible for ensuring the validity of all data processing for any selected return.

Data Editing

Due to substantial penalties for misreporting, the income and expenditure data reported on tax returns have proven to be more reliable than comparable survey data. Even so, SOI employees go to great lengths to protect against nonsampling errors, such as those due to taxpayer or data processing errors. SOI economists and statisticians develop extensive on-line tests for consistency and reliability based on the structure of the tax law and the improbability of various data combinations. Subsamples of work are independently reprocessed and compared as a further check.

Missing data problems arise, albeit infrequently (under 1 percent of the time). To handle these, missing items can sometimes be obtained through telephone or written follow-ups or through imputation. For example, an estimate can be imputed based on other data on the return (or accompanying schedules), prior-year data for the same taxpayer, or same-year data from a "statistically similar" return.

Weighting and Estimation

On the whole, the SOI approach to making aggregate (population) estimates is quite straightforward. The probability with which a return is selected for an SOI sample depends on the sampling rate prescribed for the stratum in which it is classified. Weights are computed by dividing the population count of returns filed for a given stratum by the count of sample returns for that same stratum.

In some studies, the estimates are improved by post-stratification, based on supplemental criteria or refinements of those used in the original stratification. Weights are then computed for these post-strata using refined population counts -- oftentimes with fairly computer-intensive methods.

Of over 200 million tax returns processed each year for administrative purposes, only about half a million are sampled for the SOI programs. However, since sampling rates generally increase with

increases in the size of financial amounts (for example, income or assets), the returns in the samples are, on average, disproportionately larger and more complex than those in the master files, which include the population of returns. Thus, in comparison to IRS administrative processing, which captures 100 percent of the tax returns but limited item content, SOI programs collectively represent a smaller volume, but with a proportionately higher fraction of complex returns and with much greater item content.

Publications, Products, and Services

Statistics of Income information is published both on paper and in electronic media. SOI information is published in the quarterly *SOI Bulletins* [4]; annual "complete reports" on individual [5] and corporation [6] tax returns; the annual *Corporation Source Book* [7], which provides industry and asset size data in more detail than the annual reports; and special compendiums of research and analysis [8, 9, 10]. Also, as previously noted, research articles are published in a methodology series [3].

Public-use microdata files of individuals (for which taxpayer identifiers and other means of re-identification have been removed) and certain tax-exempt organizations (whose returns are open to the public) are also available [11]. Examples of the electronic media products include the Individual Public-use File [12], as well as the *Corporation Source Book*. Also available are an electronic bulletin board [13] and a statistical information services office [14], both of which facilitate dissemination of SOI data and reports.

■ **SOI Programs**

For most of SOI's 80-year history, the main emphasis has been individual and corporation income tax information. However, growth has occurred in the nature and number of studies undertaken. Brief overviews of the major SOI studies, including work on individual, business, and international taxation, plus other special studies, are provided.

Individual Taxation Programs

Income and tax statistics from individual income tax returns have been published annually by the IRS beginning with Tax Year 1916. The content of the program is largely determined by OTA, for use in tax policy research and in tax revenue estimation. The needs of other researchers for individual income tax data are addressed on a cost-reimbursable basis.

While the individual program has historically been based on an annual cross-sectional sample of individual tax returns, a major redesign of the program has been deployed (Czajka and Walker, 1989). From discussions with OTA, it became apparent that the individual program needed to be refocused in three respects:

- Since the annual cross-sectional samples were not conducive to multi-year economic modeling for such events as sales of capital assets, the redesigned sample includes a large panel of individuals embedded within the annual cross-sectional samples.
- Because "family economic units" (reflecting households rather than individuals) are a more desirable focus of tax analysis, social security numbers of dependents, now reported on individual tax returns, are used to obtain dependents' returns, which are then combined with the parents' returns to form "tax families" (Hostetter, 1992).
- Sampling stratifiers and selection rates have been restructured to enhance the samples of returns with greater policy interest, such as those with very high or low incomes or those of the aged (Hostetter et al., 1990).

Other studies closely related to individual taxation issues are the Sales of Capital Assets (SOCA) panel studies and Information Returns studies. The SOCA studies have been conducted periodically by creating a panel file of returns to track multi-year capital asset transactions (Longton et al., 1994).

An SOI file of individual income tax returns has been linked, on a record-by-record basis, to information returns provided by employers and financial institutions (for example, Forms W-2 for wages and Forms 1099 for interest and dividends). As an offshoot of the "family economic units" study, information documents of individuals claimed as dependents on returns in this sample are being added as well (Sailer et al., 1993).

A recent *SOI Bulletin* article provided an overview of the studies and publications in the area of individual income and taxes (Sailer, 1994). In addition to the annual complete report [5], the following other reports are produced in the general area of individual taxation:

- Preliminary individual income tax data, based on an early cut-off of the sample (Gross, 1994),
- Sole proprietorships statistics, which are described in the following section on business programs (Strudler and Shiley, 1994),
- Individual income tax rates and tax shares (Nutter, 1994a),
- High-income tax returns [15], and
- Metadata systems for data products [16].

Business Taxation Programs

Most business activity in the United States is conducted by corporations, proprietorships, and partnerships, and annual SOI studies are conducted in each of these areas. While these studies are carried out independently, some recent analyses have examined the composition and dynamics of all legal forms of business (Petska and Wilson, 1994; Petska, 1993).

Corporations

The SOI corporate data are the only publicly-available source of financial information on all corporations, since other sources include only large or publicly-held corporations or those in certain regulated industries. The corporation program is rich in item content; complete income statement, balance

sheet, and tax computation information have been mainstays of the program almost since its inception. These files are the basic source of data used for estimating corporate profits for the national accounts.

As with the individual income tax studies, this program is being restructured to better meet the needs of OTA and BEA. Increased longitudinality is being designed into future studies, and enhancements are underway to compile corporate "families" by linking parent and subsidiary entities of "affiliated groups" filing consolidated returns (Hinkins et al., 1990; Moglen et al., 1987). Through the financial support of BEA, the delivery of corporation statistics has been greatly accelerated, beginning for Tax Year 1990.

As for individuals, SOI corporate data have been published annually, beginning with tax returns for 1916 [6]. In addition to the annual corporation complete report, corporate data are also published in a detailed *Source Book* [7], in the *SOI Bulletin* (Clark and Treubert, 1994), and in special analyses, such as that of the corporation "alternative minimum tax" (Treubert and Gill, 1993).

Proprietorships

Information about nonfarm proprietorship business activities is reported on Schedule C of the individual tax return. Profits from these activities are combined with income from other sources in order to compute individual "adjusted gross income." From this perspective, the proprietorship acts essentially as a conduit through which the income of the business is passed through to the business owner.

Data on proprietorships provide information on the unincorporated business sector for the national accounts. The tax return is the only annual source of financial information about these businesses. Proprietorship information is published annually in the *SOI Bulletin* (Strudler and Shiley, 1994).

Partnerships

Partnerships also serve as conduits between businesses and their owners, who are, in this case, partners. The partnership entity is not taxed directly; each partnership files an annual information return,

which includes an income statement, balance sheet (in most cases), and a schedule of allocations or distributions made to each partner. Partners report their allocated shares of income and expenses on their own tax returns. Partners are predominately, though not exclusively, individuals.

The annual SOI partnership program is important for the national accounts since it is the only source of data on these businesses. For many years, partnerships commanded only modest interest, since they were not taxed directly and, thus, had no direct effect on Federal revenue. However, the proliferation of partnerships in tax shelters in the 1980's substantially increased interest. For example, curbing deductions of partnership losses by individual partners by means of the passive loss limitations was a key provision of the 1986 Tax Reform Act (Petska, 1992; Petska and Nelson, 1990). Partnership data are also published annually in the *SOI Bulletin* (Wheeler, 1994).

International Studies

International studies are conducted annually, biennially, or periodically in two broadly-defined areas: foreign investment and activity abroad by U.S. "persons" and investment and activity in the United States by foreign "persons." Consideration is now being given to conducting more of these studies annually to meet the needs of Treasury tax policymakers.

Foreign investment and activity abroad by U.S. persons includes the following studies: corporation foreign tax credit, Controlled Foreign Corporations (of U.S. corporations), Interest Charge-Domestic International Sales Corporations, Foreign Sales Corporations, U.S. Possessions Corporations, international boycott participation, individual foreign tax credit, individual income earned abroad, and foreign trusts. Treasury uses many of these statistics in mandated reports to Congress.

Investment and activity in the United States by foreign persons include the following studies: foreign-owned U.S. corporations, foreign corporations with income derived from U.S. sources, nonresident alien income and tax withheld, nonresident alien es-

tates, U.S. partnership income of foreign partners, and sales of U.S. real property interests by foreign "persons."

The results of many of the international studies have been published in the *SOI Bulletin*. Much of the recent analyses of these data have been compiled in a compendium of studies of international income and taxes [8], while an overview of these studies was also recently published (Nutter, 1994b).

Other Special Studies

Annual, biennial, or periodic statistical programs are also conducted on tax-exempt (i.e., nonprofit) organizations, certain tax-exempt obligations, estates and personal wealth, and excise taxation.

Tax-Exempt Organizations

Annual studies are conducted of information returns filed by private foundations, nonprofit charitable and other organizations exempt under Internal Revenue Code section 501(c), exempt organizations with taxable "unrelated business income," and tax-exempt private activity bonds. Results of these studies are regularly published in the *SOI Bulletin* and have also been compiled in an historical compendium [9]. In addition, a recent article provided an overview of SOI studies of tax-exempt organizations (Skelly, 1994).

Estates and Personal Wealth

Estate tax studies are conducted annually, based both on a filing year and a specific year of death. Periodically, studies are undertaken to estimate the wealth of top (living) wealthholders by combining mortality rates with the estate data (Mallet, 1908). A long-term research project is also underway, based on estate tax filings from 1916 to the present, examining intergenerational transfers of wealth through inheritance (McCubbin, 1990). The *SOI Bulletin* is usually the primary means of disseminating the results of these studies (Johnson, 1993; Johnson and Schwartz, 1993), and an historical compendium has also recently been released [10].

SOI has also had a partnership role with the Federal Reserve Board in periodically mounting the Sur-

vey of Consumer Finances, a series of household interviews designed to estimate personal wealth (Kennickell and Woodburn, 1992).

Excise Taxes

These studies include (or have included) returns of the quarterly crude oil windfall profit tax and the environmental excise tax on certain hazardous substances -- i.e., the so-called "Superfund Tax" (Boroshok, 1993). A recent analysis of the system of Federal excise taxes has also been conducted to provide a more comprehensive and timely picture of tax receipts (Davie, 1993).

■ Limitations and Improvements

Despite many improvements, SOI systems continue to have limitations for which efforts are ongoing to address, subject to available resources. Currently, SOI is focused on several initiatives concerning these deficiencies. Five examples of these are described, followed by a sketch of some of the plans proposed to address these shortcomings. These five are: the needs for greater program timeliness, improved data consistency, better tracking of demographic changes, improved preservation of historical information, and expanded data access.

Timeliness

The fact that users never have enough current information from tax returns is an inherent weakness of the SOI program. Timeliness of SOI studies has long been a focus for improvement and one in which some successes have been achieved.

All of the major SOI studies have a sampling period that extends for one year (or more) beyond the close of the applicable accounting period to ensure the inclusion of most late filed returns. To illustrate, most Tax Year 1993 individual taxpayers filed by April 15, 1994, but significant numbers of taxpayers sought and were granted extensions until August 15, 1994. For taxpayers living abroad, further extensions were possible. Thus, to ensure inclusion of returns filed late, the sample for Tax Year 1993 was kept open throughout Calendar Year 1994.

Figure B provides a summary on the public release of SOI information. In this figure, the tax year, filing period, sampling period, and release date are all provided for most major SOI studies.

Significant efforts are being made throughout SOI to complete statistical processing within a minimum time after the close of the sampling period. What is not apparent in these summaries is the extent to which delivery dates have been improving. In addition, initiatives are underway to provide even more current information to our principal customers on a preliminary basis, by early closure of the sample periods. This has been done for the individual SOI program for many years and has recently been re-established in the corporation and the partnership programs.

Data Consistency

Problems of data consistency are of two general types -- statistical and conceptual. Despite extensive validity testing and other quality improvement efforts, inconsistent or erroneous data still escape undetected in some SOI files. Sometimes such errors are the result of "creative reporting" by taxpayers that, when weighted to produce national estimates, result in large anomalies. For whatever reason, any error that affects an estimate by a "significant" amount (where "significant" is defined by the user) is unacceptable, and work is ongoing to continue to drive the error rate down toward zero.

The ever-changing concepts of the economic and financial variables, as reflected on tax and information returns, create a multitude of difficulties in trying to examine tax data over time. Unlike other statistical agencies, where the questionnaire content can be customized to meet a statistical need, SOI's "questionnaires" are tax returns, which are developed specifically for tax compliance or other administrative purposes, rather than for statistics. Not only is there limited flexibility in changing tax return content, but the information provided on returns is often complicated, not uniformly reported, and conceptually inconsistent over time, thereby impeding multi-year economic analysis.

Figure B.--Public Release of SOI Information

| Statistics of Income Program | End of Tax Year | Close of filing period— | | Close of sampling period | Release date (on or before) |
|---|------------------|-------------------------|-------------------------|--------------------------|-----------------------------|
| | | Regular | With extensions of time | | |
| | (1) | (2) | (3) | (4) | (5) |
| Individual Income Tax Returns: | | | | | |
| 1993 | | | | | |
| Preliminary | December 1993 | April 1994 | August 1994 | September 1994 | February 1995 |
| Complete | December 1993 | April 1994 | August 1994 | December 1994 | August 1995 |
| 1994 | | | | | |
| Preliminary | December 1994 | April 1995 | August 1995 | September 1995 | February 1996 |
| Complete | December 1994 | April 1995 | August 1995 | December 1995 | August 1996 |
| Nonfarm Sole Proprietorships: | | | | | |
| 1993 | December 1993 | April 1994 | August 1994 | December 1994 | August 1995 |
| 1994 | December 1994 | April 1995 | August 1995 | December 1995 | August 1996 |
| Partnerships: | | | | | |
| 1993 | December 1993 | April 1994 | October 1994 | December 1994 | July 1995 |
| 1994 | December 1994 | April 1995 | October 1995 | December 1995 | July 1996 |
| Corporations:¹ | | | | | |
| 1992 | June 1993 | September 1993 | March 1994 | June 1994 | March 1995 |
| 1993 | June 1994 | September 1994 | March 1995 | June 1995 | March 1996 |
| 1994 | June 1995 | September 1995 | March 1996 | June 1996 | March 1997 |
| Corporation Foreign Tax Credit:² | | | | | |
| 1992 | June 1993 | September 1993 | March 1994 | June 1994 | April 1995 |
| 1993 | June 1994 | September 1994 | March 1995 | June 1995 | April 1996 |
| 1994 | June 1995 | September 1995 | March 1996 | June 1996 | April 1997 |
| Controlled Foreign Corporations:² | | | | | |
| 1992 | June 1993 | September 1993 | March 1994 | June 1994 | May 1995 |
| 1993 | June 1994 | September 1994 | March 1995 | June 1995 | May 1996 |
| 1994 | June 1995 | September 1995 | March 1996 | June 1996 | May 1997 |
| Tax-Exempt Organizations: | | | | | |
| 1991 | November 1992 | April 1993 | October 1993 | December 1993 | November 1994 |
| 1992 | November 1993 | April 1994 | October 1994 | December 1994 | August 1995 |
| 1993 | November 1994 | April 1995 | October 1995 | December 1995 | August 1996 |
| 1994 | November 1995 | April 1996 | October 1996 | December 1996 | August 1997 |
| Private Foundations: | | | | | |
| 1992 | November 1993 | April 1994 | October 1994 | December 1994 | August 1995 |
| 1993 | November 1994 | April 1995 | October 1995 | December 1995 | August 1996 |
| 1994 | November 1995 | April 1996 | October 1996 | December 1996 | August 1997 |
| Tax-Exempt Organization Unrelated Business Income: | | | | | |
| 1991 | November 1992 | April 1993 | October 1993 | December 1993 | November 1994 |
| 1992 | November 1993 | April 1994 | October 1994 | December 1994 | August 1995 |
| 1993 | November 1994 | April 1995 | October 1995 | December 1995 | August 1996 |
| 1994 | November 1995 | April 1996 | October 1996 | December 1996 | August 1997 |
| Estate Tax Returns: | | | | | |
| 1993 | (³) | (³) | (³) | December 1993 | December 1994 |
| 1994 | (³) | (³) | (³) | December 1994 | October 1995 |

¹Corporation statistics for 1992 represent accounting periods ended July 1992-June 1993. Other "years" are similarly defined.

²Dates coincide with the Corporation Statistics of Income program because these programs are based on schedules attached to the corporation income tax return.

³Estate tax return statistics are processed on a filing-year, rather than on a year-of-death basis. At a later stage (not shown here), the underlying data from several filing years are combined by year -of-death.

Improving the conceptual clarity and year-to-year consistency of the content of tax and information returns is a problem that has no easy solution. Since the tax laws have been frequently revised by the Congress, preserving year-to-year consistency can be challenging. For some data items, such as depreciation, no attempt is made to make adjustments to the basic source data provided on tax and information returns as long as they are consistent with current provisions of the tax code. In other cases, such as when the overall profitability of partnerships was redefined by a tax form change, efforts were made to re-create a consistent time series.

Tracking Demographic Changes

The redesign of the individual program at the request of OTA has underscored the need to improve the longitudinality in SOI studies. Transactions such as capital asset realizations, that can have multi-year ramifications, can best be examined by means of a panel database. As a result of this redesign, additional efforts are needed to ensure completeness as well as inter-temporal consistency.

A similar need for greater longitudinality applies to the business sector studies, where tracking large businesses can be a difficult and costly aspect of file processing. For example, ensuring the inclusion of all large corporations in the annual corporation program is a very labor-intensive effort. Often mergers and acquisitions are determined to be the source of what had initially appeared to be missing entities. Tax reforms, particularly those that have affected individual and corporate tax rates, have increased the occurrence of changes of legal form, such as switching from a corporation to a limited partnership (Petska and Wilson, 1994; Petska, 1993). However, systematic means to track such changes do not exist.

Preservation of Historical Information

Although current efforts are focused on better meeting current and future customer needs, SOI has become the "keeper" of an abundance of tax information documents in a variety of media. Much of

this information, though cumbersome to use, is irreplaceable. However, as new technologies become available, the cost of moving this information into more user-friendly formats will drop considerably. A difficult decision must be addressed as to how many current resources should be diverted from present work to safeguard this historical information.

One example might better illustrate this issue. Estate tax returns that were filed in the early part of this century had a "retention schedule" which planned for their destruction in 75 years, beginning in the early 1990's. From an archival perspective, these are an irreplaceable source of information to trace the intergenerational flows of large wealthholdings. Thus far, we have been able to delay this destruction and are attempting to capture these data for future study.

CD-ROM technologies offer great improvement in retaining data for extended periods of time compared to that of magnetic tape. SOI has in its possession literally hundreds of tape files that deteriorate in relatively short periods of time. Efforts are underway to move these files to CD-ROM, so that they will be accessible for years to come.

Public Access

Tax returns are protected by law from public scrutiny, and strict procedures govern the handling of returns and computer tape files containing such information. Even after specific identifiers (e.g., name, address, and social security number) are removed, the remaining tax return data are usually still confidential.

SOI's primary customers are authorized to receive detailed tax return (microdata) files, so computer tape files of tax return information are regularly provided to them. However, most other users of SOI data can only have access to summary tabulations. Public-use microdata files of individual tax data have been produced regularly since 1960. These files are particularly important, since they are the only source of information on high-income individuals and the only reliable source of information on property income.

Currently, access to confidential taxpayer data, even for research or analytical purposes, is available only in the following circumstances:

- ❑ **Statutory Decree** -- The Treasury, the Joint Committee on Taxation, and the Commerce Department (i.e., BEA and the Census Bureau, for certain statistical purposes) have been given access under Section 6103(f),(h), and (j) of the Internal Revenue Code (IRC) [17].
- ❑ **Federal/State Cooperative Arrangements** -- Another form of access is provision of Federal tax data to State governments for tax compliance purposes (under IRC Section 6103 (d)) [17].
- ❑ **Contractual Relationship** -- A third type of potential access to tax data is as a contractor to an organization that is allowed access. Generally, such agreements, under IRC Section 6103(n), are to allow an approved organization to contract work to a third party, so this type of release is often for computer processing or statistical support [17].
- ❑ **Intergovernmental Personnel Act Agreements** -- A fourth type of access to confidential tax data is through Intergovernmental Personnel Act (IPA) agreements. Under such arrangements, a researcher at an institution of higher learning could be given limited access to tax data for purposes that were mutually beneficial to the organization and IRS.

Making more tax microdata publicly available to researchers outside of government has been studied in SOI for some time. This has particularly been a problem for business data, where very large businesses are easy to identify. A research project to study the feasibility of public-use business files was undertaken in the early 1980's (Spruill, 1983). Various measures were considered to make these data available, while protecting taxpayer confidentiality, including purging names and other unique identifiers, rounding data items to make it harder to identify businesses, and averaging the financial data of "similar" taxpayers. While this work demonstrated

the feasibility of creating public-use data for unincorporated businesses, the cost of producing such data would be prohibitively expensive and the usefulness of the resulting data set is considered suspect. The current technological environment and availability of other public business microdata have only served to amplify these concerns [18].

■ Summary and Conclusions

The IRS is in the process of completely revamping how information is compiled for tax administration purposes. Current plans under the Tax Systems Modernization initiative are to develop a system in which:

- ❑ Taxpayers are encouraged to file returns electronically, and
- ❑ Machine-readable images are produced for returns filed on paper.

As a small part of IRS, SOI is monitoring these developments, since they will have a significant impact on statistical processing.

Meanwhile, many efforts are underway in SOI to make information from the tax system more reliable, available, and useful for policymakers, tax administrators, business planners, and the public. SOI's data processing systems have, for the most part, been modernized. Computer-assisted engineering tools are being used to develop still better systems. Sampling methods have been steadily improved, with longitudinality increasingly designed into the annual cross-sectional samples. Systems have been developed to better control the logistics of sampled returns between processing sites, and better methods of measuring processing quality and efficiency have been developed.

Current efforts in desktop publishing are contributing to increased timeliness of publications. An automated information retrieval system is being implemented that will help expedite responses to information requests. The electronic bulletin board continues to expand, and plans are underway to better link the system to the research community [13].

In short, incremental improvements are underway in virtually all of SOI's operations that should benefit SOI's many customers and users.

■ Acknowledgment

Portions of this paper have appeared in four other articles, all of which were co-authored by Fritz Scheuren, former SOI Director, and his assistance in this article is graciously acknowledged. See Scheuren and Petska, 1993; Petska and Scheuren, 1992a; Petska and Scheuren, 1992b; and Petska et al., 1992.

■ Footnotes

- [1] See *Internal Revenue Code*, Section 6108(a), "Statistical Publications and Studies, Publication or Other Disclosure of Statistics of Income."
- [2] Microsimulation modeling has become the *modus operandi* in the U.S. and many other countries for tax policy and distributional analysis. See, for example, Citro, Constance F. and Hanushek, Eric A. (eds.) (1991), *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Volume 1, Review and Recommendations*, National Academy Press.
- [3] Technical papers are published in the SOI methodology reports series, which includes this volume. See Alvey, Wendy and Jamerson, Bettye (eds.), *Statistics of Income: Turning Administrative Systems into Information Systems, 1991-1993*, Internal Revenue Service and Alvey, Wendy; Kilss, Beth; and Jamerson, Bettye (eds.), *Statistics of Income and Related Administrative Record Research -- 1981-1990*, Internal Revenue Service.
- [4] See, for example, Statistics of Income (1995), *Statistics of Income Bulletin*, Winter 1994-1995, 14, 3, Internal Revenue Service.
- [5] See, for example, Statistics of Income (1994), *Statistics of Income -- 1991, Individual Income Tax Returns*, Publication 1304, Internal Revenue Service.
- [6] See, for example, Statistics of Income (1994), *Statistics of Income -- 1991, Corporation Income Tax Returns*, Publication 16, Internal Revenue Service.
- [7] See, for example, Statistics of Income (1994), *Source Book of Statistics of Income -- 1991, Corporation Income Tax Returns*, Publication 1053, Internal Revenue Service.
- [8] See, for example, Statistics of Income (1991), *Statistics of Income -- Compendium of Studies of International Income and Taxes, 1984-1988*, Publication 1267, Internal Revenue Service.
- [9] See, for example, Statistics of Income (1993), *Statistics of Income -- Compendium of Studies of Tax-Exempt Organizations, Volume 2, 1986-1992*, Publication 1416, Internal Revenue Service.
- [10] See, for example, Statistics of Income (1994), *Statistics of Income -- Compendium of Federal Estate Tax and Personal Wealth Studies*, Publication 1773, Internal Revenue Service.
- [11] Various measures are employed to make public-use files available while protecting taxpayer confidentiality, including purging unique identifiers, rounding data items, and averaging the data of "similar" returns. For a discussion of these techniques, see Spruill, Nancy (1982), "Measures of Confidentiality," *Proceedings of the Section on Survey Research Methods, American Statistical Association*. For a description of an application of these techniques on the SOI individual income tax public-use file, see Strudler, Michael; Oh, H. Lock; and Scheuren, Fritz (1986), "Protection of Taxpayer Confidentiality with Respect to the Tax Model," *Proceedings of the Section on Sur-*

vey Research Methods, American Statistical Association.

- [12] Statistics of Income (1994), "General Description Booklet for the 1991 Individual Public-Use File," Internal Revenue Service.
- [13] For information on how to access the Electronic Bulletin Board, contact Jim Willis, Systems Administrator, at (202) 874-0277.
- [14] The Statistical Information Services office can be reached by writing to the Director, Statistics of Income CP:R:S, Internal Revenue Service, P.O. Box 2608, Washington, DC 20013-2608; by calling (202) 874-0410; or by faxing, (202) 874-0922.
- [15] See, for example, "High-Income Tax Returns, 1991," *Statistics of Income Bulletin*, Winter 1994-1995, 14, 3, Internal Revenue Service.
- [16] Metadata are essentially clarifying information about data. For example, some of the components of the metadata system include data editing instructions and sample specifications. For details, on the individual program, see David, Martin (1994) "Creating Desktop Documentation: Individual Income Tax Return Microdata, Statistics of Income," *Proceedings of the Section on Survey Research Methods, American Statistical Association* (in this volume).
- [17] See *Internal Revenue Code*, Section 6103, "Confidentiality and Disclosure of Returns and Return Information."
- [18] The Office of Management and Budget's recent *Report on Statistical Disclosure Limitation Methodology* (Subcommittee on Disclosure Limitation Methodology, Federal Committee on Statistical Methodology, 1994) provides a thorough discussion of current disclosure practices for release of Federal tabular and microdata.

■ References

- Boroshok, Sara P. (1993), "Environmental Excise Taxes, 1991," *Statistics of Income Bulletin*, Summer, 1993, 13, 1, Internal Revenue Service.
- Clark, Allison and Treubert, Patrice (1994), "Corporation Income Tax Returns, 1991," *Statistics of Income Bulletin*, Summer 1994, 14, 1, Internal Revenue Service.
- Czajka, John and Walker, Bonnye (1989), "Combining Panel and Cross-sectional Selection in an Annual Sample of Tax Returns," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Davie, Bruce (1993), "Excise Taxes, Fiscal Year 1992," *Statistics of Income Bulletin*, Fall 1993, 13, 2, Internal Revenue Service.
- Gross, Edward B. Jr. (1994), "Individual Income Tax Returns, Preliminary Data, 1992," *Statistics of Income Bulletin*, Spring 1994, 13, 4, Internal Revenue Service.
- Harte, James M. (1986), "Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Hinkins, Susan; Mulrow, Jeri; and Collins, Richard (1990), "Design and Use of an Imbedded Panel in the SOI Corporate Sample," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Hostetter, Susan (1992), "Managing Multiple Uses of Panels," *Proceedings of the Social Statistics Section, American Statistical Association*.
- Hostetter, Susan; Czajka, John; Schirm, Allan L.; and O'Connor, Karen (1990), "Choosing the Appropriate Income Classifier for Economic Tax Model-

- ling," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Johnson, Barry W. (1993), "Estate Tax Returns, 1989-1991," *Statistics of Income Bulletin*, Spring 1993, 12, 4, Internal Revenue Service.
- Johnson, Barry W. and Schwartz, Marvin (1993), "Personal Wealth, 1989," *Statistics of Income Bulletin*, Spring 1993, 12, 4, Internal Revenue Service.
- Kennickell, Arthur and Woodburn, Louise (1992), "Methodological Issues in the Estimation of the Distribution of Household Net Worth: Results from the 1989 Survey of Consumer Finances," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Longton, Barbara; Sailer, Peter; and Wilson, Janette (1994), "How Sales of Capital Assets Were Affected by the Tax Reform Act of 1986," *Proceedings of the Section on Survey Research Methods, American Statistical Association* (in this volume).
- Mallet, B. (1908), "A Method of Estimating Capital Wealth from the Estate Duty Statistics," *Journal of the Royal Statistical Society*, 71.
- McCubbin, Janet (1990), "The Intergenerational Wealth Study: Basic Estate Data, 1916-1945," *Statistics of Income Bulletin*, Spring 1990, 9, 4, Internal Revenue Service.
- Moglen, Gail; Day, Charles; and Petska, Tom (1987), "Record Linkage and Imputation Strategies in the 1982 Business Employment and Payroll Studies," *Proceedings of the Section on Survey Research Methods, American Statistical Association*
- Nutter, Sarah E. (1994a), "Individual Income Tax Rates and Tax Shares, 1990," *Statistics of Income Bulletin*, Summer 1994, 14, 1, Internal Revenue Service.
- Nutter, Sarah E. (1994b), "Statistics of Income Studies of International Income and Taxes," *Statistics of Income Bulletin*, Winter 1993-1994, 13, 3, Internal Revenue Service.
- Petska, Tom (1993), "The Effects of Tax Reform on the Structure of U.S. Business," *Proceedings of the Business and Economic Statistics Section, American Statistical Association*.
- Petska, Tom (1992), "Partnerships, Partners, and Tax Shelters after Tax Reform, 1987-1989," *Statistics of Income Bulletin*, Summer 1992, 12, 1, Internal Revenue Service.
- Petska, Tom and Nelson, Susan (1990), "Partnerships and Tax Shelters: An Analysis of the Impact of the 1986 Tax Reform," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Petska, Tom and Scheuren, Fritz (1992a), "Turning Administrative Systems into Information Systems: The Statistics of Income Program of the IRS," *Proceedings of the National Tax Association*.
- Petska, Tom and Scheuren, Fritz (1992b), "Information from Tax Returns: The Statistics of Income Program of the Internal Revenue Service, Part 2," *Business Economics*, July.
- Petska, Tom; Scheuren, Fritz; and Wilson, Bob (1992), "Information from Tax Returns: The Statistics of Income Program of the Internal Revenue Service, Part 1," *Business Economics*, April.
- Petska, Thomas B. and Wilson, Robert A. (1994), "Trends in Business Structure and Activity," *Statistics of Income Bulletin*, Spring 1994, 13, 4, Internal Revenue Service.
- Sailer, Peter (1994), "Statistics of Income Studies of Individual Income and Taxes, 1913-1992," *Statistics of Income Bulletin*, Summer 1994, 14, 1, Internal Revenue Service.

Sailer, Peter; Weber, Mike; and Yau, Ellen (1993), "How Well Can IRS Count the Population?" *Proceedings of the Government Statistics Section, American Statistical Association.*

Scheuren, Fritz and Petska, Tom (1993), "Turning Administrative Systems into Information Systems," *Journal of Official Statistics*, 9, 1.

Skelly, Daniel F. (1994), "Tax-Based Research and Data on Nonprofit Organizations, 1975-1990," *Statistics of Income Bulletin*, Summer 1994, 14, 1, Internal Revenue Service.

Spruill, Nancy (1983), "The Confidentiality and Usefulness of Masked Business Microdata," *Pro-*

ceedings of the Section on Survey Research Methods, American Statistical Association.

Strudler, Michael and Shiley, Marty (1994), "Sole Proprietorship Returns, 1992," *Statistics of Income Bulletin*, Fall 1994, 14, 2. (Annual studies of farm proprietorships were suspended after completion of the 1981 study.)

Treubert, Patrice and Gill, Amy (1993), "Effects of the Corporate Alternative Minimum Tax: An Overview of Tax Years 1990 and 1991," *Proceedings of the Business and Economic Statistics Section, American Statistical Association.*

Wheeler, Timothy D. (1994), "Partnership Returns, 1992," *Statistics of Income Bulletin*, Fall 1994, 14, 2, Internal Revenue Service. ■